On sliced Cramér metrics

William Leeb

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

Abstract

We study the family of sliced Cramér metrics, showing that they are robust to a broad class of geometric deformations. Our central results are that the sliced Cramér distance between a function and its deformation may be bounded by certain natural measures of the deformation's displacement, multiplied by the function's mean mixed norm. These results extend to sliced Cramér distances between tomographic projections. We also remark on the effect of convolution on the sliced Cramér metrics. We compare these properties of sliced Cramér metrics to similar properties satisfied by Wasserstein distances. In addition, we study computationally efficient Fourier-based discretizations of the Cramér and sliced Cramér distances in 1D and 2D, and prove that they are robust to heteroscedastic noise. The results are illustrated by numerical experiments.

1 Introduction

This paper is concerned with properties of the sliced Cramér metrics, a family of distances that extend the classical univariate Cramér metrics [12, 56, 67, 79] to functions in $\mathbb{R}^d$ via the "slicing" operation. Sliced Cramér metrics have been the subject of investigation in a number of recent works in machine learning and image processing [6, 39, 29, 61], where they are contrasted with sliced Wasserstein distances. Our main results describe the robustness of these distances to deformations of the input functions. The question of robustness is natural when considering which metric to use for a given application, as one would often like to use a metric that is insensitive to small perturbations of the input data. Prior work has explored the robustness of Wasserstein-type metrics, including the sliced Wasserstein distances, to certain geometric deformations. In the present work, we will show that the sliced Cramér metrics exhibit similar properties. In addition, we also study other properties of practical interest, namely the sliced Cramér metrics' behavior under convolutions and in the presence of additive heteroscedastic noise.

1.1 Geometric deformations

It is often desirable for a metric to be robust to a specified class of deformations of the inputs. For example, in image processing, one may seek a metric that is insensitive to small translations, rotations, changes of scale, or perturbations in the parameters that generate the image. While there are many classes of deformations of practical interest, in this paper, we consider the family of push-forward deformations of a function: if Φ is a C^1 bijection mapping into the domain of a function f, it induces the deformation $f_\Phi(x) = f(\Phi(x))|\nabla\Phi(x)|$ of f. Such deformations preserve f's integral and L^1 norm. Other works have considered metric stability under different classes of deformations, such as those that preserve f's L^∞ norm [42, 3]; however, integral-preserving deformations are more natural for understanding sliced Cramér metrics, which are often used to compare probability measures.

In prior work [37, 54, 61], it is shown that certain metrics are robust to the action of deformations, in the sense that the distance between a function and its deformation under Φ may be bounded above by a suitable measure of the "size" of Φ. One possible definition of deformation size is the maximum displacement, $\max_x |x - \Phi(x)|$. In this paper, we consider a family of measures of displacement size, defined in Section 2.1, which includes the maximum displacement as a special case. Proposition 2.1 in Section 2.1 relates the different measures of deformation size.

1.2 Wasserstein and sliced Wasserstein distances

Prior works on metric robustness have focused primarily on the family of Wasserstein distances, defined between probability distributions. Informally, the p-Wasserstein distance between two probability distributions f and g is equal to the minimal cost of transforming f into g by rearranging the mass, where the cost of moving mass from x to y is proportional to $|x - y|^p$ [70, 71]; we review the precise definition of Wasserstein distances in Section 2.3. The Wasserstein distances are popular metrics in a range of machine learning and statistical applications [50, 51, 59, 9, 52, 58, 38, 8, 55, 46, 10]. Sliced Wasserstein distances are defined by averaging the 1D Wasserstein distances over all 1D tomographic projections of the functions [6, 39, 29, 61].

It is known that Wasserstein-type distances exhibit several kinds of geometric robustness, as explored in the prior works [37, 54, 61]. Specifically, [54, 61] show robustness to rigid deformations (translations and rotations) for comparing 2D tomographic projections of a 3D volume; this suggests that these distances – and other distances satisfying similar properties – are a natural choice for studying tomographic images, such as those arising from single particle cryogenic electron microscopy (cryo-EM) [64, 7, 15], as well as other scientific problems for which the measurement modality only permits observing projections of an object [13, 48, 26, 27, 65, 11].

In Section 2, specifically Theorem 2.2 and Theorem 2.3, we state fairly general robustness properties of Wasserstein and sliced Wasserstein distances. Though we have not seen these results published previously, they follow straightforwardly from prior work, and further help elucidate why Wasserstein distances are often a useful class of distances. The fact that Wasserstein-type distances are robust to deformations is not surprising, as robustness is essentially "baked into" the optimization problem that defines the Wasserstein distance. Indeed, from the Monge formulation of Wasserstein distance it is tautologically true that the p-Wasserstein distance between a probability density f and its deformation f_Φ is bounded by the maximum displacement of Φ. It is therefore natural to ask whether other families of metrics exhibit similar robustness properties, even when such robustness is not an intrinsic part of their definition. This is one of the motivations behind our present work.

1.3 Cramér and sliced Cramér metrics

The focus of this work is the family of sliced Cramér metrics. We will recall the precise definition of these metrics in Section 2.6: briefly, the p-Cramér distance between univariate functions f and g is the L^p distance between the Volterra operator (the indefinite integral operator) applied to f and g, and the sliced Cramér distance between multivariate functions f and g is defined by averaging the Cramér distances between all one-dimensional tomographic projections of f and g.

The Cramér and sliced Cramér metrics have been proposed as alternatives to the Wasserstein and sliced Wasserstein distances in certain machine learning environments [6, 39, 29]. Despite this, it appears that the geometric properties of the sliced Cramér metrics (or even the Cramér metrics themselves) have not been thoroughly explored. Furthermore, whereas Wasserstein and sliced Wasserstein distances are defined between probability distributions, certain applications require distances between functions with both positive and negative values (such as when observations are corrupted by additive noise). Because sliced Cramér distances are defined between any two functions, not just probability distributions, and because the two families of metrics have similar definitions (in fact, the sliced 1-Cramér and sliced 1-Wasserstein metrics are identical), it is natural to consider using sliced Cramér distances in these settings. It therefore becomes of particular interest to understand whether they share similar robustness properties as Wasserstein-type distances.

We will show that, like Wasserstein and sliced Wasserstein distances, the sliced Cramér distances are provably robust to deformations. In fact, the robustness bounds we prove for sliced Cramér metrics are in some sense stronger than those for Wasserstein distances: more precisely, when $p > 1$, the sliced p-Cramér metric is bounded by a concave, non-linear function of the deformation's maximum displacement, whereas the corresponding bound on Wasserstein grows linearly with the maximum displacement. That is, while the bounds are similar for small displacements, the bounds for the sliced Cramér metrics are stronger for large displacements. Of course, whether or not this is an advantage will depend on the specific application; nevertheless, it is of interest to understand the differences in behavior between these metrics.

1.4 Noise and convolutions

In addition to showing their robustness to deformations, we will remark on the behavior of sliced Cramér metrics under convolutions of the input functions. In typical scientific applications, each observation will be a convolution of the underlying physical object with a filter that arises from the measurement apparatus. It is therefore of interest to understand how a metric changes when the inputs are each convolved by a common function. Since convolution is a smoothing operation, one expects that the convolved functions will be closer to each other than the clean functions. Indeed, this is the case for L^p distances (due to Young's convolution inequality), and is also true for Wasserstein and sliced Wasserstein distances. We make the simple observation that the same is true for sliced Cramér distances. This result essentially appears in the paper [74]; though the statement there is restricted to comparing probability distributions, it is straightforward to extend it to the case of general functions.

We will also study computationally efficient, Fourier-based discretizations of the 1D Cramér metric and 2D sliced Cramér metric; the latter is based on the discretization described in [61]. We will show that this discretization is robust to additive, heteroscedastic Gaussian noise: that is, when applied to samples from a signal-plus-noise model, the estimated distance converges to the distance between the signals only, and filters out the noise. This is a useful property for a metric to have (though one that is by no means unique to the sliced Cramér metrics – there are many methods for filtering out noise). The result also highlights a difference between the sliced Cramér and Wasserstein metrics, alluded to earlier in Section 1.3: because Wasserstein-type distances are defined between probability distributions, it is not always clear how to define them for noisy samples, whose values may be positive or negative and whose sums may not be equal. By contrast, sliced Cramér metrics are a priori defined between functions with mixed signs and unequal integrals.

1.5 Mean mixed norms

The main results in this paper bound the sliced Cramér distance between a function f and its deformation f_Φ. Of course, these bounds necessarily depend on both the function f and the deformation Φ. We have already alluded to the dependence on Φ in Section 1.1. The dependence on f is through its so-called mean mixed norm, defined by averaging f's mixed L^1-L^p norms; the precise definition may be found in Section 2.2. While they may appear complicated at first glance, the use of the function's mean mixed norm, as opposed to the more simpy defined L^p norm, permits bounds that do not explicitly depend on the size of f's support. We also observe (see Remark 1 in Section 2.2 below) that compactly supported functions in L^p necessarily have finite mean mixed norm.

1.6 Notation and conventions

Throughout the paper, we will assume familiarity with basic concepts of measure and integration, e.g. at the level of [19] or [25], and the rudiments of probability theory. In this section we briefly review several definitions and introduce the notational conventions that we will use.

1.6.1 Lebesgue norms

Throughout, we denote by $\|f\|_{L^p} = \left(\int_{\mathbb{R}^d} |f(x)|^p dx\right)^{1/p}$ the Lebesgue p-norm of a function $f : \mathbb{R}^d \to \mathbb{R}$ (with the obvious modification when $p = \infty$), and denote the standard inner product as $\langle f, g\rangle = \int_{\mathbb{R}^d} f(x)g(x)dx$. We also define the p-norm $\|x\|_p$ for vectors x in $\mathbb{R}^d$ as $\|x\|_p = \left(\sum_{j=1}^d |x_j|^p\right)^{1/p}$ (again, with the obvious modification when $p = \infty$). When convenient, we will also denote the 2-norm of a vector x in $\mathbb{R}^d$ by $|x|$. We also denote the inner product between two vectors x and y in $\mathbb{R}^d$ by $\langle x, y\rangle = \sum_{j=1}^d x_j y_j$.

1.6.2 Absolute continuity

For a given interval $(a, b) \subset \mathbb{R}$, we denote by $\mathcal{A}_0 = \mathcal{A}_0(a, b)$ the set of absolutely continuous functions G on (a, b) satisfying $G(b) = 0$; note that such functions may be written in the form $G(x) = -\int_x^b g(t)\, dt$ where $g = G'$ almost everywhere.

1.6.3 The Fourier transform

We denote the Fourier transform of a function $f : \mathbb{R}^d \to \mathbb{R}$ in L^1 by $\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x)e^{-2\pi \mathrm{i}\langle x,\xi\rangle} dx$. With this convention, under mild conditions f may be recovered using the inverse Fourier transform $f(x) = \int_{\mathbb{R}^d} \widehat{f}(\xi)e^{2\pi \mathrm{i}\langle x,\xi\rangle} d\xi$.

1.6.4 Tomographic projections and the Radon transform

Let $\mathcal{U} = \left(u^{(1)}, \ldots, u^{(r)}\right) \in \mathbb{S}^{d-1} \times \cdots \times \mathbb{S}^{d-1}$ (where $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ is the $(d-1)$-dimensional unit sphere) denote an ordered collection of r unit vectors in $\mathbb{R}^d$. Let $u^{(r+1)}, \ldots, u^{(d)}$ denote $d-r$ orthonormal vectors that are orthogonal to $u_1, \ldots, u_r$. We define the tomographic projection $\mathcal{P}_{\mathcal{U}}$ onto $\operatorname{span}\{u^{(1)}, \ldots, u^{(r)}\}$ by

$$(\mathcal{P}_{\mathcal{U}} f)(t_1, \ldots, t_r) = \int_{\mathbb{R}^{d-r}} f(t_1 u^{(1)} + \cdots + t_r u^{(r)} + s_1 u^{(r+1)} + \cdots + s_{d-r} u^{(d)})\, ds. \tag{1}$$

When $r = 1$, we denote the tomographic projection of f onto the span of a unit vector u by $\mathcal{P}_u f$. Note that in this case, the *Radon transform* $\mathcal{R}f : \mathbb{R} \times \mathbb{S}^{d-1}$ of the function f is defined by $(\mathcal{R}f)(t, u) = (\mathcal{P}_u f)(t)$. For more background on these transforms, see, for example, the references [48, 26]. A standard result that we will use is the *Fourier slice theorem*: $\widehat{(\mathcal{P}_u f)}(\xi) = \widehat{f}(\xi u)$, for any unit vector $u \in \mathbb{R}^d$ and real number ξ.

1.6.5 Push-forwards

If $\Omega \subset \mathbb{R}^d$ is a (non-empty) open set, μ is a finite, signed measure on Ω, and $\Psi : \Omega \to \mathbb{R}^d$ is a measurable function, we denote by $\Psi_\sharp \mu$ the push-forward measure, $(\Psi_\sharp \mu)(E) = \mu(\Psi^{-1}(E))$; e.g. see [51]. Note that $(\Psi_\sharp \mu)(\Psi(\Omega)) = \mu(\Psi^{-1}(\Psi(\Omega))) = \mu(\Omega)$. When μ is induced from a function f supported on Ω, i.e. $\mu(E) = \int_E f(x)dx$, and Ψ is a diffeomorphism between Ω and $\Psi(\Omega)$ with inverse $\Phi = \Psi^{-1}$, then $\Psi_\sharp \mu$ has density $f(\Phi(x))|\det(\nabla \Phi(x))|$. We will write $(\Psi_\sharp f)(x) = (\Phi_\sharp^{-1} f)(x) = f(\Phi(x))|\det(\nabla \Phi(x))|$, or $f_\Phi(x) = f(\Phi(x))|\det(\nabla \Phi(x))|$ for short.

1.7 Outline of the remainder of the paper

The remainder of the paper is structured as follows:

1. Section 2 introduces key definitions and background material. Theorem 2.2 and Theorem 2.3 establish general robustnes properties of Wasserstein and sliced Wasserstein distances, respectively; these results appear to be new, though they follow easily from prior literature.

2. Section 3 states the main theorems on the robustness of sliced Cramér distances to geometric deformations. Theorem 3.1 and Corollary 3.2 give bounds under quite general conditions. Sharper bounds are proved for several special cases of interest in Section 3.2, including rotations, translations, and dilations. In addition, Theorem 3.13 describes the behavior of sliced Cramér metrics under convolutions.

3. Section 4 analyzes Fourier-based approximations to the Cramér distances and the 2D sliced Cramér distances, the latter with respect to the uniform measure over $\mathbb{S}^1$. Theorems 4.1 and 4.10 show that these discretizations are robust to additive heteroscedastic Gaussian noise.

4. Section 5 shows the results of several numerical experiments illustrating the theoretical results, including comparisons between the sliced Cramér, sliced Wasserstein, and Lebesgue distances.

5. Section 6 concludes the paper, providing a summary and topics for future research.

2 Background definitions and theory

This section introduces key definitions and background results that will be referred to in the rest of the paper. We draw particular attention to Theorem 2.2 and Theorem 2.3 on the robustness of Wasserstein and sliced Wasserstein distances, which we have not seen published in the literature (but which follow straightforwardly from previously published results).

2.1 Deformations and displacement

Given an open, non-empty $\Omega \subset \mathbb{R}^d$ and a C^1, invertible Ψ defined on Ω with inverse $\Phi = \Psi^{-1}$, we will call the push-forward $f_\Phi(x) = f(\Phi(x))|\det(\nabla \Phi(x))|$ a *deformation of* f, and will also call the mappings Φ and Ψ themselves deformations.

We define the *maximum displacement* of the deformation Φ by

$$\varepsilon_\infty(\Phi) = \max_{x \in \Psi(\Omega)} |x - \Phi(x)|. \tag{2}$$

For example, if u is a fixed unit vector, then the function $\Psi(x) = x + \epsilon u$ has $\varepsilon_\infty(\Psi) = \epsilon$.

If u is a unit vector and Ψ is a C^1, 1-to-1 mapping defined on Ω, we define the maximum displacement of Ψ along u to be

$$\varepsilon(\Psi, u) = \max_{x \in \Omega} |\langle x - \Psi(x), u\rangle|. \tag{3}$$

Note that the maximum displacement can be written as

$$\varepsilon_\infty(\Psi) = \max_{u \in \mathbb{S}^{d-1}} \varepsilon(\Psi, u). \tag{4}$$

For a given probability distribution η over $\mathbb{S}^{d-1}$, and a value $1 \le p < \infty$, we define the *mean displacement* of Ψ as

$$\varepsilon_{\eta,p}(\Psi) = \left(\int_{\mathbb{S}^{d-1}} \varepsilon(\Psi, u)^p \, d\eta(u) \right)^{1/p}. \tag{5}$$

Note that if $\Phi = \Psi^{-1}$, then for all u,

$$\begin{aligned}
\varepsilon(\Psi, u) &= \max_{x \in \Omega} |\langle x - \Psi(x), u\rangle| \\
&= \max_{y \in \Psi(\Omega)} |\langle \Phi(y) - \Psi(\Phi(y)), u\rangle| \\
&= \max_{y \in \Psi(\Omega)} |\langle \Phi(y) - y, u\rangle| \\
&= \varepsilon(\Phi, u).
\end{aligned} \tag{6}$$

It then follows immediately that $\varepsilon_\infty(\Phi) = \varepsilon_\infty(\Psi)$ and $\varepsilon_{\eta,p}(\Phi) = \varepsilon_{\eta,p}(\Psi)$. When η is the uniform measure, we will denote $\varepsilon_{\eta,p}(\Psi)$ by $\varepsilon_p(\Psi)$. Also, when $d = 1$, all measures of distortion are identical, and we will denote their common value by $\varepsilon(\Psi)$.

Clearly, $\varepsilon_{\eta,p}(\Psi) \le \varepsilon_\infty(\Psi)$ for all p. The following result shows what is lost when the inequality is reversed:

Proposition 2.1. *Let $x^* = \arg\max_x |x - \Psi(x)|$, and let $u^* = (x^* - \Psi(x^*))/|x^* - \Psi(x^*)|$. Then*

$$\varepsilon_{\eta,p}(\Psi) \ge \varepsilon_\infty(\Psi) \cdot \left(\int_{S^{d-1}} |\langle u^*, u\rangle|^p \, d\eta(u) \right)^{1/p}. \tag{7}$$

Proof. By definition, $\varepsilon_\infty(\Psi) = |x^* - \Psi(x^*)|$. For all unit vectors u,

$$\varepsilon(\Psi, u) = \max_x |\langle x - \Psi(x), u\rangle| \ge |\langle x^* - \Psi(x^*), u\rangle| = |x^* - \Psi(x^*)| \cdot |\langle u^*, u\rangle| = \varepsilon_\infty(\Psi) \cdot |\langle u^*, u\rangle|. \tag{8}$$

The result then follows from averaging over u.

□

For example, if $d = 2$ and η is the uniform measure over $\mathbb{S}^1$,

$$\left(\int_{S^1} |\langle u^*, u\rangle|^p \, d\eta(u) \right)^{1/p} = \left(\frac{2}{\pi} \int_0^{\pi/2} \cos(\theta)^p \, d\theta \right)^{1/p} = \left(\frac{\Gamma(p/2 + 1/2)}{\Gamma(p/2+1)\sqrt{\pi}} \right)^{1/p}. \tag{9}$$

2.2 Mean mixed norms

Fix a probability measure η over the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$. For a function f on $\mathbb{R}^d$, we define the *mean mixed norm*

$$\|f\|_{M^{p,r}_\eta} = \left(\int_{\mathbb{S}^{d-1}} \|\mathcal{P}_u(|f|)\|^r_{L^p(\mathbb{R})} d\eta(u) \right)^{1/r} \tag{10}$$

for any $1 \leq p \leq \infty$ and $1 \leq r < \infty$; and

$$\|f\|_{M^{p,\infty}} = \|f\|_{M_\eta^{p,\infty}} = \operatorname*{ess\,sup}_{u \in \mathbb{S}^{d-1}} \|\mathcal{P}_u(|f|)\|_{L^p(\mathbb{R})}. \tag{11}$$

When $p = r$, we will define $\|f\|_{M_\eta^p} \equiv \|f\|_{M_\eta^{p,p}}$.

For a fixed u, $\|\mathcal{P}_u(|f|)\|_{L^p(\mathbb{R})}$ is an example of a so-called *mixed norm* [28], namely, the L^p norm of the L^1 norm of f. The mean mixed norm is then obtained by averaging this mixed norm over the choice of L^p variable.

Remark 1. If f is supported on a bounded open set and $f \in L^p$, then $\|f\|_{M_\eta^{p,r}} < \infty$. More precisely, if f is supported on a ball of radius $R > 0$, then, $\|f\|_{M_\eta^{p,r}} \leq (2R)^{(d-1)(p-1)/p}\|f\|_{L^p}$.

Remark 2. If η is the uniform measure, or if $r = \infty$, then $\|f\|_{M_\eta^{p,r}}$ is rotationally-invariant.

2.3 Wasserstein distances

If f and g are probability densities on a subset $\Omega \subset \mathbb{R}^d$, their *p-Wasserstein distance* $\mathrm{W}_p(f, g)$ (also known as the Kantorovich distance) is defined as

$$\mathrm{W}_p(f, g) = \min_{\Pi \in \mathcal{M}(f,g)} \left(\int_\Omega \int_\Omega |x - y|^p d\Pi(x, y) \right)^{1/p}, \tag{12}$$

where $\mathcal{M}(f, g)$ denotes the space of all probability measures on $\Omega \times \Omega$ with marginals equal to f and g, respectively [70, 71]. That is, $\Pi \in \mathcal{M}(f, g)$ if for all measurable $E \subset \Omega$,

$$\Pi(E \times \Omega) = \int_E f(x)dx, \tag{13}$$

and

$$\Pi(\Omega \times E) = \int_E g(y)dy. \tag{14}$$

Informally, $\mathrm{W}_p(f, g)$ is the minimal cost of rearranging a unit of mass with distribution f into one with distribution G, where the cost of moving mass between locations x and y is $|x - y|^p$. The distance $\mathrm{W}_1(f, g)$ is also known as the *Earth Mover's Distance (EMD)* between the probability measures f and g [70, 71]. The Wasserstein distances and their variants have been widely used in statistics, machine learning, image processing, and related areas [50, 51, 59, 9, 52, 58, 38, 8, 55, 46, 10].

The Wasserstein distance is a relaxation of the Monge distance, defined by

$$\mathrm{M}_p(f, g) = \min_{\Phi \in \mathcal{T}(f,g)} \left(\int_\Omega |x - \Phi(x)|^p f(x)\, dx \right)^{1/p} \tag{15}$$

where $\mathcal{T}(f, g)$ contains those functions $\Phi : \Omega \to \Omega$ such that $\int_E g(x)dx = \int_{\Phi^{-1}(E)} f(x)dx$, that is, which which push f onto g. Indeed, any Φ in $\mathcal{T}(f, g)$ induces a measure Π_Φ in $\mathcal{M}(f, g)$, with

$$\int_\Omega \int_\Omega |x - y|^p d\Pi_\Phi(x, y) = \int_\Omega |x - \Phi(x)|^p f(x)\, dx, \tag{16}$$

and hence $\mathrm{W}_p(f, g) \leq \mathrm{M}_p(f, g)$. (In fact, when $\mathrm{M}_p(f, g)$ is finite, equality holds; see [59].) Consequently, if Φ is a smooth bijection on Ω and $f_\Phi(x) = f(\Phi(x))|\det(\nabla\Phi(x))|$, then Φ is contained in $\mathcal{T}(f, f_\Phi)$, and so

$$\mathrm{W}_p(f, f_\Phi) \leq \mathrm{M}_p(f, f_\Phi) \leq \left(\int_\Omega |x - \Phi(x)|^p f(x)\, dx \right)^{1/p} \leq \varepsilon_\infty(\Phi) \left(\int_\Omega f(x)\, dx \right)^{1/p} = \varepsilon_\infty(\Phi). \tag{17}$$

In fact, a more general robustness result may be easily shown, which we state now. The proof is nearly identical to that found in [54].

Theorem 2.2. *Suppose f is a probability density supported on a bounded, open set $\Omega \subset \mathbb{R}^D$, and let $\Phi : \Omega \to \Omega$ be a smooth bijection. Let $\mathcal{Q}$ be a tomographic projection operator onto a d-dimensional subspace, $d \leq D$. Then for all $p \geq 1$,*

$$\mathrm{W}_p(\mathcal{Q}f, \mathcal{Q}f_\Phi) \leq \varepsilon_\infty(\Phi). \tag{18}$$

Proof. An identical proof to that of Lemma 1 in [54] shows that $\mathrm{W}_p(\mathcal{Q}f, \mathcal{Q}f_\Phi) \le \mathrm{W}_p(f, f_\Phi)$ (note that the left side refers to transportation in $\mathbb{R}^d$, and the right side to $\mathbb{R}^D$). The bound then follows from (17).

□

It is well-known that Wasserstein distances in 1D take a on a particularly simple form. Denote by $\mathcal{V}$ the Volterra operator [23] on $L^1([a, b])$, defined by

$$(\mathcal{V}f)(x) = \int_a^x f(t)dt. \tag{19}$$

Then when $d = 1$, it is known [59] that $\mathrm{W}_p(f, g)$ may be written as follows:

$$\mathrm{W}_p(f, g) = \|(\mathcal{V}f)^{-1} - (\mathcal{V}g)^{-1}\|_{L^p}. \tag{20}$$

Here, $(\mathcal{V}f)^{-1}$ denotes the functional inverse of $\mathcal{V}f$, defined as

$$(\mathcal{V}f)^{-1}(x) = \inf\{t \in [a, b] \ : \ (\mathcal{V}f)(t) \ge x\}. \tag{21}$$

When $p = 1$, it is also true that $\mathrm{W}_1(f, g) = \|\mathcal{V}f - \mathcal{V}g\|_{L^1}$.

2.4 Sliced metrics

Given a metric $\mathrm{D}(f, g)$ between univariate functions, a value $p \ge 1$, and a probability density η over the unit sphere in $\mathbb{R}^d$, one can define a corresponding *sliced* metric $\mathrm{SD}_{\eta,p}(f, g)$ defined between functions f and g of d variables, as follows:

$$\mathrm{SD}_{\eta,p}(f, g) = \left(\int_{\mathbb{S}^{d-1}} \mathrm{D}(\mathcal{P}_u f, \mathcal{P}_u g)^p d\eta(u) \right)^{1/p}. \tag{22}$$

That is, $\mathrm{SD}_{\eta,p}(f, g)$ is obtained by averaging the distances between the one-dimensional projections of f and g over all directions.

Taking D to be the p-Wasserstein distance W_p, we denote by $\mathrm{SW}_{\eta,p}$ the corresponding sliced Wasserstein distance. Sliced Wasserstein distances have been the subject of considerable research activity in recent years [52, 9, 31, 32, 14, 49, 61]. The work [61] proves that sliced Wasserstein distances are robust to rotations and translations, and also describes a fast discretization, which we will make use of in the present work.

We prove an analogue of Theorem 2.2 for sliced Wasserstein distances:

Theorem 2.3. *Suppose f is a probability density supported on a bounded, open set $\Omega \subset \mathbb{R}^D$, and let $\Phi : \Omega \to \Omega$ be a smooth bijection. Let $\mathcal{Q}$ be a tomographic projection operator onto d-dimensional subspace, $d \le D$. Then for all $p \ge 1$,*

$$\mathrm{SW}_{\eta,p}(\mathcal{Q}f, \mathcal{Q}f_\Phi) \le \varepsilon_\infty(\Phi). \tag{23}$$

Proof. Without loss of generality, suppose $\mathcal{Q}$ projects onto the first d coordinates. Let $u \in \mathbb{S}^{d-1}$. It is easy to see (and will be shown later, in the proof of Corollary 3.2 in Section 3) that $(\mathcal{P}_u \mathcal{Q}f)(t) = (\mathcal{P}_{(u,0)} f)(t)$, which is the tomographic projection of f onto the span of $(u, 0) \in \mathbb{R}^d \times \mathbb{R}^{D-d}$. Consequently, from Theorem 2.2,

$$\mathrm{W}_p(\mathcal{P}_u \mathcal{Q}f, \mathcal{P}_u \mathcal{Q}f_\Phi) = \mathrm{W}_p(\mathcal{P}_{(u,0)} f, \mathcal{P}_{(u,0)} f_\Phi) \le \varepsilon_\infty(\Phi). \tag{24}$$

The result now follows by averaging over u.

□

2.5 Wasserstein distances and convolution

In signal and image processing applications, one typically observes signals/images that have been convolved with a filter induced from the measurement process. It is therefore of interest to understand how metrics behave when their inputs are convolved by a common function. For Wasserstein distances, we have the following result:

Theorem 2.4. *Suppose f, g and w are probability densities on $\mathbb{R}^d$. Then for all $p \ge 1$,*

$$\mathrm{W}_p(f * w, g * w) \le \mathrm{W}_p(f, g). \tag{25}$$

This property is referred to by Zolotarev as *regularity* of the metric [78, 77, 79].[1] The result is certainly known and is easy to prove, though we had some difficulty finding a reference for a proof of the precise statement; for the reader's convenience we provide a proof here.

Proof of Theorem 2.4. By Kantorovich duality (see, e.g., Chapter 5 in [71]),

$$\mathrm{W}_p(f,g)^p = \sup_{(\varphi,\psi)\in\mathcal{F}_p} \left\{ \int f(x)\varphi(x)\,dx + \int g(y)\psi(y)\,dy \right\}, \tag{26}$$

where $\mathcal{F}_p$ contains all pairs (φ,ψ) of integrable functions φ and ψ satisfying $\varphi(x)+\psi(y) \le |x-y|^p$. Take any such φ and ψ. Letting $\widetilde{w}(z) = w(-z)$, we have

$$\int (f*w)(x)\varphi(x)\,dx = \int (\varphi * \widetilde{w})(z) f(z)\,dz, \tag{27}$$

and

$$\int (g*w)(y)\psi(y)\,dy = \int (\psi * \widetilde{w})(z) g(z)\,dz. \tag{28}$$

For any x and y we have

$$(\varphi*\widetilde{w})(x) + (\psi*\widetilde{w})(y) = \int (\varphi(x-z)+\psi(y-z))\widetilde{w}(z)\,dz \le |x-y|^p \int \widetilde{w}(z)\,dz = |x-y|^p. \tag{29}$$

Therefore, $(\varphi*\widetilde{w}, \psi*\widetilde{w}) \in \mathcal{F}_p$, and

$$\int (f*w)(x)\varphi(x)\,dx + \int (g*w)(y)\psi(y)\,dy = \int (\varphi*\widetilde{w})(z)f(z)\,dz + \int (\psi*\widetilde{w})(z)g(z)\,dz \le \mathrm{W}_p(f,g), \tag{30}$$

and so taking the supremum over all $(\varphi,\psi)\in\mathcal{F}_p$ proves the result.

□

Corollary 2.5. *Suppose f, g and w are probability densities on $\mathbb{R}^d$, and let η be a probability density over $\mathbb{S}^{d-1}$. Then for all $p \ge 1$,*

$$\mathrm{SW}_p(f*w, g*w) \le \mathrm{SW}_p(f,g). \tag{31}$$

Proof. It is straightforward to see (and will be shown in Section 3.3) that for any unit vector u in $\mathbb{R}^d$, $\mathcal{P}_u(f*w) = (\mathcal{P}_u f)*(\mathcal{P}_u w)$. Then from the $d=1$ case of Theorem 2.4,

$$\mathrm{W}_p(\mathcal{P}_u(f*w), \mathcal{P}_u(g*w)) = \mathrm{W}_p((\mathcal{P}_u f)*(\mathcal{P}_u w), (\mathcal{P}_u g)*(\mathcal{P}_u w)) \le \mathrm{W}_p(\mathcal{P}_u f, \mathcal{P}_u g). \tag{32}$$

Averaging over all $u \in \mathbb{S}^{d-1}$ then proves the result.

□

Remark 3. Young's convolutional inequality (e.g. see Chapter 8 in [19]) states that the same property holds for the ordinary L^p distances on $\mathbb{R}^d$, namely, for f and g in L^p and w in L^1, $\|f*w - g*w\|_{L^p} \le \|w\|_{L^1}\|f-g\|_{L^p}$.

2.6 Cramér and sliced Cramér metrics

Let f be in $L^1(a,b)$. For any value $1 \le p \le \infty$, we will refer to $\|f\|_{V^p} \equiv \|\mathcal{V}f\|_{L^p}$ as the *Volterra p-norm* of f. Note that, because $\mathcal{V}f$ is in $L^\infty(a,b)$, the Volterra p-norm of f is finite for any function f in $L^1(a,b)$. We denote by $\mathrm{C}_p(f,g) = \|f-g\|_{V^p}$ the *p-Cramér metric* (or *p-Cramér distance*) between functions f and g, named after Harald Cramér [12, 56, 67].[2]

Remark 4. While the Cramér metrics are often used to compare probability distributions, they are well-defined for any f and g in $L^1(a,b)$. Our analysis of the Cramér metrics in this paper is not restricted to probability densities.

[1] An equivalent statement of regularity for a metric D between random variables is that for all random variables X, Y and Z, where Z is independent of X and Y, $D(X+Z, Y+Z) \le D(X,Y)$.

[2] In [79], Zolotarev refers to the p-Cramér distance more simply as the "L_p-metric".

Remark 5. When f and g are probability measures, the ∞-Cramér metric is also known as the Kolmogorov Metric between f and g [22]. The Kolmogorov Metric arises in the context of goodness-of-fit testing in statistics [21]. The 1-Cramér metric is equal to the 1-Wasserstein distance, or *Earth Mover's Distance*, between the probability distributions f and g described in Section 2.3. The 2-Cramér metric is the 1D *energy distance* [56].

The following result provides a dual formulation of the Volterra p-norm (and hence of the p-Cramér metric) that will be useful in our subsequent analysis. It essentially appears as Theorem 1 in [41]; because of its key role in this paper, we provide a self-contained proof for the reader's convenience.

Proposition 2.6. *Let $1 \leq p \leq \infty$ and let q be the conjugate exponent: $1/p + 1/q = 1$. Then for any function f in $L^1(a, b)$,*

$$\|f\|_{V^p} = \sup_{G \in \mathcal{A}_0 : \|G'\|_{L^q} \leq 1} \langle f, G \rangle. \tag{33}$$

Proof. First, note that the adjoint transform $\mathcal{V}^*$ is given by

$$(\mathcal{V}^* f)(x) = \int_x^b f(t) dt. \tag{34}$$

This operator satisfies

$$\langle \mathcal{V} f, g \rangle = \langle f, \mathcal{V}^* g \rangle \tag{35}$$

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

By duality of L^p and L^q, we have:

$$\|f\|_{V^p} = \|\mathcal{V} f\|_{L^p} = \sup_{g : \|g\|_{L^q} \leq 1} \int_a^b (\mathcal{V} f)(x) g(x) dx = \sup_{g : \|g\|_{L^q} \leq 1} \langle \mathcal{V} f, g \rangle = \sup_{g : \|g\|_{L^q} \leq 1} \langle f, \mathcal{V}^* g \rangle. \tag{36}$$

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:

$$\|f\|_{V^p} = \sup_{g : \|g\|_{L^q} \leq 1} \langle f, \mathcal{V}^* g \rangle = \sup_{G \in \mathcal{A}_0 : \|G'\|_{L^q} \leq 1} \langle f, G \rangle, \tag{37}$$

which completes the proof.

□

Following the framework from Section 2.4, given a probability measure η over the unit sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$, for all $1 \leq p < \infty$ we define *sliced p-Cramér metric* between $f, g : \mathbb{R}^d \to \mathbb{R}$ as

$$\mathrm{SC}_{\eta,p}(f, g) = \left(\int_{\mathbb{S}^{d-1}} \mathrm{C}_p(\mathcal{P}_u f, \mathcal{P}_u g)^p \, d\eta(u) \right)^{1/p}, \tag{38}$$

and

$$\mathrm{SC}_\infty(f, g) = \mathrm{SC}_{\eta,\infty}(f, g) = \sup_{u \in \mathbb{S}^{d-1}} \mathrm{C}_\infty(\mathcal{P}_u f, \mathcal{P}_u g). \tag{39}$$

Sliced Cramér metrics have garnered attention in recent years as metrics for comparing probability measures in machine learning applications [47, 30, 6, 39, 29]. In Section 3, we will study the geometric properties of sliced Cramér metrics. In Section 4 we will study efficient, Fourier-based discretizations for the 1D and 2D distances (based on those in [61]) between functions with equal integrals, and prove their robustness to additive heteroscedastic noise.

For a function f of two variables, we define its *sliced Volterra norm* by

$$\|f\|_{SV^p_\eta} = \left(\int_{\mathbb{S}^{d-1}} \|\mathcal{P}_u f\|_{L^p}^p \, d\eta(u) \right)^{1/p} \tag{40}$$

when $1 \leq p < \infty$, and

$$\|f\|_{SV^\infty} = \|f\|_{SV^\infty_\eta} = \sup_{u \in \mathbb{S}^{d-1}} \|\mathcal{P}_u f\|_{L^\infty}. \tag{41}$$

Then for two functions f and g, $\mathrm{SC}_{\eta,p}(f, g) = \|f - g\|_{SV^p_\eta}$.

When η is the uniform measure over $\mathbb{S}^{d-1}$, we will denote the sliced Cramér metric more simply as $\mathrm{SC}_p(f, g)$, and the sliced Volterra norm more simply as $\|f\|_{SV^p}$.

3 Properties of sliced Cramér metrics

In this section, we will prove that the sliced Cramér metrics defined in Section 2.6 are robust to deformations of the input functions; more precisely, that the distance between a function and its deformation is bounded by a monotonic function of the deformations's maximum displacement. These results are similar in spirit to the bounds for Wasserstein and sliced Wasserstein distances found in Theorem 2.2 and Theorem 2.3, respectively. Bounds for general deformations are provided in Theorem 3.1 and Corollary 3.2, in Section 3.1. Sharper bounds are then derived for more specific deformations in Section 3.2. Section 3.3 analyzes the behavior of the sliced Cramér distances under convolution of the inputs, stating a bound analagous to Theorem 2.4 and Corollary 2.5.

3.1 Robustness to deformations

We start with a general result that quantifies sliced Cramér metrics' robustness to deformations.

Theorem 3.1. *Let $1 \leq p \leq \infty$. Let A and B be non-empty, bounded, open sets in $\mathbb{R}^d$, f be in $L^p(A)$, $\Phi : B \to A$ be a C^1 bijection, and $f_\Phi(x) = f(\Phi(x))|\det(\nabla\Phi(x))|$ on B, and 0 elsewhere. Then for any probability measure η over $\mathbb{S}^{d-1}$,*

$$\mathrm{SC}_{\eta,p}(f, f_\Phi) \leq 2^{(p-1)/p} \cdot \|f\|_{M^p_\eta} \cdot \varepsilon_\infty(\Phi) \tag{42}$$

$$\mathrm{SC}_{\eta,p}(f, f_\Phi) \leq 2^{(p-1)/p} \cdot \|f\|_{M^{p,\infty}} \cdot \varepsilon_{\eta,p}(\Phi), \tag{43}$$

and

$$\mathrm{SC}_{\eta,p}(f, f_\Phi) \leq \|f\|_{L^1} \cdot \varepsilon_{\eta,1}(\Phi)^{1/p}. \tag{44}$$

Remark 6. For two functions f and g, the same upper bounds hold trivially for $|\mathrm{SC}_{\eta,p}(f,g) - \mathrm{SC}_{\eta,p}(f_\Phi, g)|$. That is, the theorem bounds the difference between the distances when an input function is replaced by a deformation.

Remark 7. Because $f = (f_\Phi)_{\Phi^{-1}}$ and $\varepsilon(\Phi, u) = \varepsilon(\Phi^{-1}, u)$, one can switch the roles of f and f_Φ, and thereby replace the term $\|f\|_{M^p_\eta}$ by $\min\{\|f\|_{M^p_\eta}, \|f_\Phi\|_{M^p_\eta}\}$, and replace $\|f\|_{M^{p,\infty}_\eta}$ by $\min\{\|f\|_{M^{p,\infty}_\eta}, \|f_\Phi\|_{M^{p,\infty}_\eta}\}$, for a sharper bound.

Theorem 3.1 is easily extended to comparing tomographic projections of a function and its deformation. This is of interest when measuring the distance between two 2D projections of a 3D volume, such as in the analysis of images in cryo-electron microscopy.

Corollary 3.2. *Let $1 \leq p \leq \infty$. Let A and B be non-empty, bounded, open sets in $\mathbb{R}^D$, f be in $L^p(A)$, $\Phi : B \to A$ be a C^1 bijection, and $f_\Phi(x) = f(\Phi(x))|\det(\nabla\Phi(x))|$ on B, and 0 elsewhere. For $d \leq D$, let $\mathcal{Q}$ denote the tomographic projection operator onto a d-dimensional subspace of $\mathbb{R}^D$. Then for any probability measure η over $\mathbb{S}^{d-1}$,*

$$\mathrm{SC}_{\eta,p}(\mathcal{Q}f, \mathcal{Q}f_\Phi) \leq 2^{(p-1)/p} \cdot \|\mathcal{Q}(|f|)\|_{M^p_\eta} \cdot \varepsilon_\infty(\Phi), \tag{45}$$

and

$$\mathrm{SC}_{\eta,p}(\mathcal{Q}f, \mathcal{Q}f_\Phi) \leq \|f\|_{L^1} \cdot \varepsilon_\infty(\Phi)^{1/p}. \tag{46}$$

Proof of Corollary 3.2. Without loss of generality, suppose $\mathcal{Q}$ projects onto the first d coordinates, that is,

$$(\mathcal{Q}f)(x) = \int_{\mathbb{R}^{D-d}} f(x, y) dy. \tag{47}$$

Let u be a unit vector in $\mathbb{R}^d$, and let $u^{(2)}, \ldots, u^{(d)}$ denote any orthonormal vectors completing the basis, so that for h on $\mathbb{R}^d$,

$$(\mathcal{P}_u h)(t) = \int_{\mathbb{R}^{d-1}} h(tu + s_2 u^{(2)} + \cdots + s_d u^{(d)})\, ds, \tag{48}$$

and so

$$
\begin{aligned}
(\mathcal{P}_u(\mathcal{Q}h))(t) &= \int_{\mathbb{R}^{d-1}} (\mathcal{Q}h)(tu + s_2 u^{(2)} + \cdots + s_d u^{(d)})\, ds \\
&= \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}^{D-d}} h(tu + s_2 u^{(2)} + \cdots + s_d u^{(d)}, y)\, dy\, ds \\
&= (\mathcal{P}_{(u,0)} h)(t),
\end{aligned} \tag{49}
$$

which is the tomographic projection of h onto the span of $(u, 0) \in \mathbb{R}^d \times \mathbb{R}^{D-d}$.

Denote by $\widetilde{\eta}$ the distribution over the unit sphere $\mathbb{S}^{D-1}$, supported on $\widetilde{\mathbb{S}}^{d-1} \equiv \{(u,0) \in \mathbb{R}^d \times \mathbb{R}^{D-d} : |u| = 1\}$ and defined by $d\widetilde{\eta}((u,0)) = d\eta(u)$ for $u \in \mathbb{R}^d$. We have

$$
\begin{aligned}
\mathrm{SC}_{\widetilde{\eta},p}(f, f_\Phi)^p &= \int_{\mathbb{S}^{D-1}} \mathrm{C}_p(\mathcal{P}_v f, \mathcal{P}_v f_\Phi)^p\, d\widetilde{\eta}(v) \\
&= \int_{\mathbb{S}^{d-1}} \mathrm{C}_p(\mathcal{P}_{(u,0)} f, \mathcal{P}_{(u,0)} f_\Phi)^p\, d\eta(u) \\
&= \int_{\mathbb{S}^{d-1}} \mathrm{C}_p(\mathcal{P}_u \mathcal{Q} f, \mathcal{P}_u \mathcal{Q} f_\Phi)^p\, d\eta(u) \\
&= \mathrm{SC}_{\eta,p}(\mathcal{Q}f, \mathcal{Q}f_\Phi)^p.
\end{aligned} \tag{50}
$$

Furthermore,

$$
\begin{aligned}
\|f\|_{M^p_{\widetilde{\eta}}}^p &= \int_{\mathbb{S}^{D-1}} \|\mathcal{P}_v(|f|)\|_{L^p}^p\, d\widetilde{\eta}(v) \\
&= \int_{\mathbb{S}^{d-1}} \|\mathcal{P}_{(u,0)}(|f|)\|_{L^p}^p\, d\eta(u) \\
&= \int_{\mathbb{S}^{d-1}} \|\mathcal{P}_u \mathcal{Q}(|f|)\|_{L^p}^p\, d\eta(u) \\
&= \|\mathcal{Q}(|f|)\|_{M^p_\eta}^p.
\end{aligned} \tag{51}
$$

The inequality (45) then follows by applying (42) in Theorem 3.1 with the measure $\widetilde{\eta}$. The bound (46) follows from (44) and the fact that $\varepsilon_{\widetilde{\eta},1}(\Phi) \le \varepsilon_\infty(\Phi)$.

□

The proof of Theorem 3.1 is immediate from the following lemma:

Lemma 3.3. *Let $1 \le p \le \infty$. Let A and B be non-empty, bounded, open sets in $\mathbb{R}^d$, f be in $L^p(A)$, $\Phi : B \to A$ be a C^1 bijection and $f_\Phi(x) = f(\Phi(x))|\det(\nabla\Phi(x))|$ on B, and 0 elsewhere. Then for any $u \in \mathbb{S}^{d-1}$,*

$$
\mathrm{C}_p(\mathcal{P}_u f, \mathcal{P}_u f_\Phi) \le \min\left\{ 2^{(p-1)/p} \cdot \|\mathcal{P}_u(|f|)\|_{L^p} \cdot \varepsilon(\Phi, u),\ \|f\|_{L^1} \cdot \varepsilon(\Phi, u)^{1/p} \right\}. \tag{52}
$$

Theorem 3.1 follows easily from averaging each side over u.

Proof of Lemma 3.3. Without loss of generality, suppose $u = e_1 = (1, 0, \ldots, 0)$; then

$$
\varepsilon(\Psi, u) = \max_{(x,y) \in \mathbb{R} \times \mathbb{R}^{d-1}} |x - \psi_1(x, y)|. \tag{53}
$$

For brevity, if $h : \mathbb{R}^d \to \mathbb{R}$ is a function of d variables, let $\mathcal{P}h = \mathcal{P}_{e_1} h$. That is,

$$
(\mathcal{P}h)(x) = \int_{\mathbb{R}^{d-1}} h(x, y) dy. \tag{54}
$$

For $(x, y) \in \mathbb{R} \times \mathbb{R}^{d-1}$, let $I_{(x,y)}$ be the interval $[x, \psi_1(x, y)]$ when $x \le \psi_1(x, y)$, and $[\psi_1(x, y), x]$ when $x > \psi_1(x, y)$; and let $\chi(x, y, t)$ be 1 if $t \in I_{(x,y)}$, and 0 otherwise; that is, $\chi(x, y, t) = 1$ if either $x \le t \le \psi_1(x, y)$ or $\psi_1(x, y) \le t \le x$.

Step 1. We will show that for all $(x, y) \in \mathbb{R} \times \mathbb{R}^{d-1}$,

$$\int \chi(x, y, t)\, dt \le \varepsilon(\Psi, u), \tag{55}$$

and for all t,

$$\int \sup_y \chi(x, y, t)\, dx \le 2\varepsilon(\Psi, u). \tag{56}$$

For the first inequality, for fixed (x, y), suppose without loss of generality that $x \le \psi_1(x, y)$. Then $\chi(x, y, t) = 1$ if and only if $x \le t \le \psi_1(x, y)$, and so

$$\int \chi(x, y, t)\, dt = |\psi_1(x, y) - x| \le \varepsilon(\Psi, u), \tag{57}$$

which is (55).

For the second inequality: for any x and t, $\sup_y \chi(x, y, t) = 1$ if and only if there exists a vector y such that either $x \le t \le \psi_1(x, y)$ or $\psi_1(x, y) \le t \le x$. In this case, since $|x - \psi_1(x, y)| \le \varepsilon(\Psi, u)$, we must also have $|x - t| \le \varepsilon(\Psi, u)$, and so x lies in the interval $[t - \varepsilon(\Psi, u), t + \varepsilon(\Psi, u)]$ of length $2\varepsilon(\Psi, u)$; hence

$$\int \sup_y \chi(x, y, t)\, dx \le 2\varepsilon(\Psi, u), \tag{58}$$

which is (56).

Step 2. We will prove that

$$\mathrm{C}_p(\mathcal{P}_u f, \mathcal{P}_u f_\Phi) \le 2^{(p-1)/p} \cdot \|\mathcal{P}_u(|f|)\|_{L^p} \cdot \varepsilon(\Phi, u). \tag{59}$$

Let $G \in \mathcal{A}_0$, with derivative $g = G'$ satisfying $\|g\|_{L^q} \le 1$. Performing the change of variables $w = \Phi(x, y)$ gives

$$\begin{aligned}
\int_{\mathbb{R}} G(x)(\mathcal{P} f_\Phi)(x)\, dx &= \int_{\mathbb{R}} G(x) \int_{y:(x,y)\in B} f(\Phi(x, y))|\det(\nabla\Phi(x, y))|\, dy\, dx \\
&= \int_B G(x) f(\Phi(x, y))|\det(\nabla\Phi(x, y))|\, dy\, dx \\
&= \int_A G(\psi_1(w)) f(w)\, dw \\
&= \int_{\mathbb{R}} \int_{y:(x,y)\in A} G(\psi_1(x, y)) f(x, y)\, dy\, dx
\end{aligned} \tag{60}$$

and similarly,

$$\int_{\mathbb{R}} G(x)(\mathcal{P} f)(x)\, dx = \int_{\mathbb{R}} \int_{y:(x,y)\in A} G(x) f(x, y)\, dy\, dx. \tag{61}$$

Combining (60) and (61), and applying Hölder's inequality,

$$\begin{aligned}
\int_{\mathbb{R}} G(x)((\mathcal{P} f)(x) - (\mathcal{P} f_\Phi)(x))\, dx &= \int \int [G(x) - G(\psi_1(x, y))] f(x, y)\, dy\, dx \\
&\le \int \left(\sup_y |G(x) - G(\psi_1(x, y))|\right) \left(\int_{\mathbb{R}^{d-1}} |f(x, y)|\, dy\right) dx \\
&= \int \left(\sup_y |G(x) - G(\psi_1(x, y))|\right) \mathcal{P}(|f|)(x) dx \\
&\le \|\mathcal{P}(|f|)\|_{L^p} \cdot \left\|\sup_y |G(x) - G(\psi_1(x, y))|\right\|_{L^q(dx)},
\end{aligned} \tag{62}$$

where $1/p + 1/q = 1$.

We will show that

$$\left\| \sup_y |G(x) - G(\psi_1(x,y))| \right\|_{L^q(dx)} \le 2^{(p-1)/p}\, \varepsilon(\Psi, u), \tag{63}$$

which yields (59) by applying Proposition 2.6. We have

$$|G(x) - G(\psi_1(x,y))| = \left| \int_x^{\psi_1(x,y)} g(t)dt \right| = \left| \int g(t)\chi(x,y,t)dt \right|. \tag{64}$$

Suppose temporarily that $1 < p, q < \infty$. Applying Hölder's inequality and using (55) and (56), we get

$$\begin{aligned}
\int \left(\sup_y |G(x) - G(\psi_1(x,y))| \right)^q dx &= \int \sup_y |G(x) - G(\psi_1(x,y))|^q dx \\
&= \int \sup_y \left| \int g(t)\chi(x,y,t)dt \right|^q dx \\
&\le \int \sup_y \left(\int |g(t)|^q \chi(x,y,t)dt \right) \left(\int \chi(x,y,t)\, dt \right)^{q/p} dx \\
&\le \varepsilon(\Psi,u)^{q/p} \int \sup_y \int |g(t)|^q \chi(x,y,t)\, dt\, dx \\
&\le \varepsilon(\Psi,u)^{q/p} \int \int |g(t)|^q \sup_y \chi(x,y,t)\, dt\, dx \\
&= \varepsilon(\Psi,u)^{q/p} \int |g(t)|^q \int \sup_y \chi(x,y,t)\, dx\, dt \\
&\le 2\varepsilon(\Psi,u)^{q/p+1} \int |g(t)|^q\, dt,
\end{aligned} \tag{65}$$

and so taking the q-th root we get the bound

$$\left\| \sup_y |G(x) - G(\psi_1(x,y))| \right\|_{L^q(dx)} \le 2^{1/q} \varepsilon(\Psi,u)^{1/p+1/q} \|g\|_{L^q} \le 2^{(p-1)/p} \varepsilon(\Psi,u). \tag{66}$$

Now suppose $p = 1$ and $q = \infty$. Then from (55),

$$\begin{aligned}
\left\| \sup_y |G(x) - G(\psi_1(x,y))| \right\|_{L^\infty(dx)} &= \sup_{x,y} |G(x) - G(\psi_1(x,y))| \\
&= \sup_{x,y} \left| \int g(t)\chi(x,y,t)dt \right| \\
&\le \|g\|_{L^\infty} \sup_{x,y} \int \chi(x,y,t)\, dt \\
&\le \|g\|_{L^\infty} \varepsilon(\Psi,u) \\
&\le \varepsilon(\Psi,u).
\end{aligned} \tag{67}$$

Finally, suppose $p = \infty$ and $q = 1$. Then from (56),

$$\begin{aligned}\left\|\sup_y |G(x) - G(\psi_1(x,y))|\right\|_{L^1(dx)} &= \int \sup_y |G(x) - G(\psi_1(x,y))|\, dx \\ &= \int \sup_y \left| \int g(t)\chi(x,y,t) dt \right| dx \\ &\leq \int \int |g(t)| \sup_y \chi(x,y,t)\, dt\, dx \\ &= \int |g(t)| \int \sup_y \chi(x,y,t)\, dx\, dt \\ &\leq \|g\|_{L^1} \sup_t \int \sup_y \chi(x,y,t)\, dx \\ &\leq \|g\|_{L^1} 2\varepsilon(\Psi, u) \\ &= 2\varepsilon(\Psi, u). \end{aligned} \tag{68}$$

This completes the proof of (63), and hence proves (59).

Step 3. To conclude the proof, we will prove the inequality

$$\mathrm{C}_p(\mathcal{P}_u f, \mathcal{P}_u f_\Phi) \leq \|f\|_{L^1} \cdot \varepsilon(\Psi, u)^{1/p}. \tag{69}$$

Let $G \in \mathcal{A}_0$, with derivative $g = G'$ satisfying $\|g\|_{L^q} \leq 1$. From (62), taking $p = 1$ and $q = \infty$,

$$\int_{\mathbb{R}} G(x)((\mathcal{P}f)(x) - (\mathcal{P}f_\Phi)(x))\, dx \leq \|f\|_{L^1} \sup_{x,y} |G(x) - G(\psi_1(x,y))|, \tag{70}$$

and so by using Proposition 2.6, it is enough to show that for all (x, y),

$$|G(x) - G(\psi_1(x,y))| \leq \varepsilon(\Psi, u)^{1/p}. \tag{71}$$

Hölder's inequality yields

$$\begin{aligned}|G(x) - G(\psi_1(x,y))| &= \left| \int g(t)\chi(x,y,t) dt \right| \\ &\leq \|g\|_{L^q} \left(\int \chi(x,y,t)^p\, dt \right)^{1/p} \\ &\leq \left(\int \chi(x,y,t)\, dt \right)^{1/p} \\ &\leq \varepsilon(\Psi, u)^{1/p}, \end{aligned} \tag{72}$$

where the last inequality follows from (55). This completes the proof.

□

Next, we will consider special cases for which quantitatively tighter bounds can be shown.

3.2 Sharper bounds in special cases

In this section, we consider specific classes of deformations, and prove sharper bounds than (42) and (43) from Theorem 3.1.

3.2.1 Rotations in 2D

We consider the case where $A = B = \mathbb{D} \subset \mathbb{R}^2$, the open unit disc centered at $(0, 0)$. If Φ is a rotation around the origin by angle θ, then the corresponding maximum displacement is $\varepsilon(\Phi) = 2\sin(\theta/2)$, and so the bound (42) in Theorem 3.1 is

$$\mathrm{SC}_{\eta,p}(f, f_\Phi) \le 2^{(p-1)/p} \cdot \|f\|_{M_\eta^p} \cdot 2\sin(\theta/2). \tag{73}$$

(We do not consider the bound (43), as for this choice of Φ it is never stronger than (42).)

We can prove a sharper estimate:

Theorem 3.4. *Let $1 \le p \le \infty$, and $f : \mathbb{R}^2 \to \mathbb{R}$ be in $L^p(\mathbb{D})$. Suppose $0 \le \theta < \pi$, and define f_θ by*

$$f_\theta(x, y) = f(x\cos(\theta) + y\sin(\theta), y\cos(\theta) - x\sin(\theta)). \tag{74}$$

Then for any probability distribution η over $\mathbb{S}^{d-1}$,

$$\mathrm{SC}_{\eta,p}(f, f_\theta) \le \|f\|_{M_\eta^p} \cdot \Delta_p(\theta), \tag{75}$$

where

$$\Delta_p(\theta) = \begin{cases} 2\sin(\theta/2) \cdot (2\cos(\theta/2))^{(p-1)/p}, & \textit{if } 0 \le \theta < \pi/2 \\ 2\sin(\theta/2)^{1/p}, & \textit{if } \pi/2 \le \theta < \pi \end{cases}. \tag{76}$$

The result follows from the following lemma:

Lemma 3.5. *Using the notation from the statement of Theorem 3.4, if u is any unit vector in $\mathbb{R}^2$, then*

$$\mathrm{C}_p(\mathcal{P}_u f, \mathcal{P}_u f_\theta) \le \|\mathcal{P}_u(|f|)\|_{L^p} \cdot \Delta_p(\theta). \tag{77}$$

Theorem 3.4 follows immediately by taking the p-th power and averaging over all u.

Proof of Lemma 3.5. Without loss of generality, suppose $u = (1, 0)$. An identical proof to that of Lemma 3.3 may be applied by replacing the bound $\sup_t \int \sup_y \chi(x, y, t)dx \le 2\varepsilon(\Phi)$ from (56) with the bound

$$\int \sup_y \chi(x, y, t)dx \le \begin{cases} 2\sin(\theta), & \text{if } 0 \le \theta < \pi/2 \\ 2, & \text{if } \pi/2 \le \theta < \pi \end{cases} \tag{78}$$

for all $|t| < 1$. Indeed, when $0 \le \theta < \pi/2$, we can then replace (63) with the upper bound

$$\begin{aligned} \left\| \sup_y |G(x) - G(\psi_1(x, y))| \right\|_{L^q(dx)} &\le (2\sin(\theta/2))^{1/p} \cdot (2\sin(\theta))^{1/q} \\ &= (2\sin(\theta/2))^{1/p} \cdot (4\sin(\theta/2)\cos(\theta/2))^{1/q} \\ &= 2\sin(\theta/2) \cdot (2\cos(\theta/2))^{1/q} \\ &= 2\sin(\theta/2) \cdot (2\cos(\theta/2))^{(p-1)/p} \end{aligned} \tag{79}$$

whereas when $\pi/2 \le \theta < \pi$ the bound becomes

$$\left\| \sup_y |G(x) - G(\psi_1(x, y))| \right\|_{L^q(dx)} \le (2\sin(\theta/2))^{1/p} \cdot 2^{1/q} = 2\sin(\theta/2)^{1/p}. \tag{80}$$

The bound $\int \sup_y \chi(x, y, t)dx \le 2$ is immediate, since the integrand is bounded by 1 and the integral is over $|x| \le 1$. Hence it remains to show that

$$\int \sup_y \chi(x, y, t)dx \le 2\sin(\theta) \tag{81}$$

whenever $0 \le \theta < \pi/2$.

Let $c = \cos(\theta)$ and $s = \sin(\theta)$. Note that in this case, $c \geq 0$ and $s \geq 0$; and the rotation $\Phi(x, y) = (cx+sy, cy-sx)$, with inverse $\Psi(x, y) = (cx - sy, cy + sx)$.

Take $|t| < 1$, and suppose, without loss of generality, that $0 \leq t \leq 1$. It is enough to show

$$\int \sup_y \chi(x, y, t)\, dx \leq \begin{cases} 2s\sqrt{1-t^2}, & \text{if } 0 \leq t < c \\ 1 - ct + s\sqrt{1-t^2}, & \text{if } c \leq t \leq 1 \end{cases}. \tag{82}$$

Indeed, the right side of (82) is decreasing in t and so is maximized when $t = 0$, which yields the desired bound (81).

We now show (82). For $0 \leq t \leq 1$, let S_t denote the set of all x, $|x| \leq 1$, satisfying $\sup_y \chi(x, y, t) = 1$. Then $\int \sup_y \chi(x, y, t)\, dx = |S_t|$.

Lemma 3.6. *Let* $0 \leq t \leq 1$.

1. *Suppose* $t \leq x$. *Then* $x \in S_t$ *if and only if* $cx - s\sqrt{1-x^2} \leq t$.
2. *Suppose* $x \leq t$. *Then* $x \in S_t$ *if and only if* $t \leq cx + s\sqrt{1-x^2}$.

Proof. First suppose that $t \leq x$. If $x \in S_t$, then there exists y with $\chi(x, y, t) = 1$, that is, for which $(x, y) \in \mathbb{D}$ and $\psi_1(x, y) = cx - sy \leq t \leq x$. Since $cx - sy$ only gets smaller as y grows, we can always take $y = \sqrt{1-x^2}$. The converse is immediate.

Next suppose that $x \leq t$. If $x \in S_t$, then there exists y for which $(x, y) \in \mathbb{D}$ and $x \leq t \leq cx - sy$. Since $cx - sy$ only gets bigger as y shrinks, we can always take $y = -\sqrt{1-x^2}$. Again, the converse is immediate.

□

Lemma 3.7. *Let* $0 \leq t \leq 1$. *Then for all* $x \in S_t$, $x \geq ct - s\sqrt{1-t^2}$.

Proof. We will break this into two cases, depending on whether $x \leq t$ or $x > t$.

Case 1. $x \leq t$. By Lemma 3.6, $cx + s\sqrt{1-x^2} \geq t$. Therefore, $s\sqrt{1-x^2} \geq t - cx$ and since we assume $x \leq t$, $t - cx \geq 0$; therefore, squaring each side gives

$$s^2 - s^2x^2 \geq t^2 + c^2x^2 - 2ctx, \tag{83}$$

and hence

$$x^2 - 2ctx + t^2 - s^2 \leq 0, \tag{84}$$

and since the roots of the polynomial are $x = ct \pm s\sqrt{1-t^2}$, in particular it follows that $x \geq ct - s\sqrt{1-t^2}$.

Case 2. $x > t$. In this case, the result follows immediately from the inequality $ct - s\sqrt{1-t^2} \leq ct \leq t < x$.

□

Lemma 3.8. *Suppose that* $0 \leq t \leq c$. *Then all* $x \in S_t$ *satisfy* $x \leq ct + s\sqrt{1-t^2}$.

Proof. First, note that the assumption $0 \leq t \leq c$ implies that $t \leq ct + s\sqrt{1-t^2}$. Indeed, the latter inequality is equivalent to $t(1-c) \leq s\sqrt{1-t^2}$, and which is equivalent to $t^2 + t^2c^2 - 2ct^2 \leq s^2 - t^2s^2$, which is equivalent to $2t^2(1-c) \leq s^2 = 1 - c^2 = (1-c)(1+c)$, which is equivalent to $2t^2 \leq 1 + c$, which is immediate from $0 \leq t \leq c \leq 1$.

Now, let $x \in S_t$. If $x \leq t$, then $x \leq ct + s\sqrt{1-t^2}$ as well, by what we have just shown. Therefore, we can suppose instead that $x > t$. By Lemma 3.6, $cx - s\sqrt{1-x^2} \leq t$. Note that $cx - s\sqrt{1-x^2}$ is an increasing function of x. Hence, if $x > t$ satisfies

$$cx - s\sqrt{1-x^2} = t, \tag{85}$$

then for all $x' > x$, we have $x' > t$ and $cx' - s\sqrt{1-(x')^2} > t$, and hence by Lemma 3.6 $x' \notin S_t$. For $x > t$ satisfying (85), we have

$$\begin{aligned} & cx - t = s\sqrt{1-x^2} \\ \implies & c^2x^2 + t^2 - 2ctx = s^2 - s^2x^2 \\ \implies & x^2 - 2ctx + t^2 - s^2 = 0 \\ \implies & x = ct \pm s\sqrt{1-t^2} \\ \implies & x = ct + s\sqrt{1-t^2}, \end{aligned} \tag{86}$$

where the last implication is because $x > t$. This completes the proof.

$\square$

From Lemmas 3.7 and 3.8, when $0 \le t < c$, all $x \in S_t$ lie between $ct - s\sqrt{1-t^2}$ and $ct + s\sqrt{1-t^2}$; hence $|S_t| \le 2s\sqrt{1-t^2}$. On the other hand, when $t \ge c$, by Lemma 3.7 all $x \in S_t$ must lie between $ct - s\sqrt{1-t^2}$ and 1; hence $|S_t| \le 1 - ct + s\sqrt{1-t^2}$. This concludes the proof of (82), and hence of Lemma 3.5.

$\square$

3.2.2 Monotonically increasing deformations in 1D

When $d = 1$, the bounds (42) and (43) in Theorem 3.1 become

$$\mathrm{C}_p(f, f_\Phi) \le 2^{(p-1)/p} \cdot \|f\|_{L^p} \cdot \varepsilon(\Phi). \tag{87}$$

Now we show that the factor of $2^{(p-1)/p}$ can be removed when the deformation Φ is monotonically increasing.

Theorem 3.9. *Let $1 \le p \le \infty$. Let I and J be non-empty, bounded, open intervals in $\mathbb{R}$, f be in $L^p(I)$, $\Phi : J \to I$ be a C^1 bijection with inverse Ψ and $\Phi'(x) > 0$ on J, and $f_\Phi(x) = f(\Phi(x))\Phi'(x))$ on J, and 0 elsewhere. Then*

$$\mathrm{C}_p(f, f_\Phi) \le \|f\|_{L^p} \cdot \varepsilon(\Phi). \tag{88}$$

Remark 8. To see that the monotonocity of Φ is required for this sharper bound, consider the following example. Fix $\eta > \delta > 0$, and let f be defined by

$$f(x) = \begin{cases} 1, & \text{if } -\eta \le x < 0, \\ -1, & \text{if } 0 \le x \le \eta, \\ 0, & \text{otherwise.} \end{cases} \tag{89}$$

Let $\Phi : [-\delta, \delta] \to [-\eta, \eta]$ be defined by $\Phi(x) = -(\eta/\delta)x$. Then

$$f_\Phi(x) = \begin{cases} -\eta/\delta, & \text{if } -\delta \le x < 0, \\ \eta/\delta, & \text{if } 0 \le x \le \delta, \\ 0, & \text{otherwise.} \end{cases} \tag{90}$$

Then it is straightforward to verify that $\varepsilon(\Phi) = \eta + \delta$, $\|f\|_{L^\infty} = 1$, and $C_\infty(f, f_\Phi) = 2\eta$. By taking $\delta \to 0$, we see that the bound $C_\infty(f, f_\Phi) \le 2\|f\|_{L^\infty}\varepsilon(\Phi)$ is tight.

Proof of Theorem 3.9. Let I_x be the interval $[x, \Psi(x)]$ if $x \le \Psi(x)$, and $[\Psi(x), x]$ if $\Psi(x) \le x$. Let $\chi(x,t)$ be 1 if $t \in I_x$, and 0 otherwise. Then an identical proof to that of Lemma 3.3 may be applied if we show that

$$\sup_t \int \chi(x,t)dx \le \varepsilon(\Phi), \tag{91}$$

in place of the bound (56).

Take any $t \in I$, and 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 $|x^* - \Psi(x^*)| \le \varepsilon(\Phi)$, it follows that $|t - x^*| \le \varepsilon(\Phi)$ 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,

$$\int \chi(x,t)dx \le \int_{x^*}^{t} 1dx = |t - x^*| \le \varepsilon(\Phi). \tag{92}$$

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

$\square$

3.2.3 Translations

We consider the case where $\Phi(x) = x + v$, where v is a fixed vector. For simplicity, we only describe the case where η is the uniform measure, though the results easily generalize. In this case, the bounds (42) and (43) from Theorem 3.1 are, respectively,

$$\mathrm{SC}_p(f, f_\Phi) \leq 2^{(p-1)/p} \cdot \|f\|_{M^p} \cdot |v| \tag{93}$$

and

$$\mathrm{SC}_p(f, f_\Phi) \leq 2^{(p-1)/p} \cdot K_p \cdot \|f\|_{M^{p,\infty}} \cdot |v|, \tag{94}$$

where

$$K_p = \left(\int_{\mathbb{S}^{d-1}} |u_1|^p \, du\right)^{1/p}. \tag{95}$$

We show that the factor $2^{(p-1)/p}$ can be removed:

Theorem 3.10. *Let $1 \leq p \leq \infty$. Let A be a non-empty, bounded, open set in $\mathbb{R}^d$, and f be in $L^p(A)$. For a fixed vector v, let $f_v(x) = f(x + v)$. Then*

$$\mathrm{SC}_p(f, f_v) \leq \|f\|_{M^p} \cdot |v| \tag{96}$$

and

$$\mathrm{SC}_p(f, f_v) \leq K_p \cdot \|f\|_{M^{p,\infty}} \cdot |v|, \tag{97}$$

where K_p is defined in (95).

Remark 9. When $p = 1$, $\mathrm{SC}_1(f, f_v) = \mathrm{SW}_1(f, f_v)$, and so the bound (97) when $p = 1$ and $d = 2$ matches the $p = 1$ case of Theorem 2 of [61].

Proof of Theorem 3.10. By a straightforward calculation, for any $u \in \mathbb{S}^{d-1}$, $(\mathcal{P}_u f_v)(t) = (\mathcal{P}_u f)(t + \langle v, u\rangle)$. Consequently, by Theorem 3.9,

$$\mathrm{C}_p(\mathcal{P}_u f, \mathcal{P}_u f_v) \leq \|\mathcal{P}_u f\|_{L^p} \cdot |\langle v, u\rangle|. \tag{98}$$

Then

$$\mathrm{SC}_p(f, f_v)^p \leq \int_{\mathbb{S}^{d-1}} \|\mathcal{P}_u f\|_{L^p}^p \cdot |\langle v, u\rangle|^p \, du \leq |v|^p \cdot \int_{\mathbb{S}^{d-1}} \|\mathcal{P}_u f\|_{L^p}^p \, du \leq \|f\|_{M^p} \cdot |v|^p, \tag{99}$$

which is (96); and

$$\mathrm{SC}_p(f, f_v)^p \leq \sup_{u \in \mathbb{S}^{d-1}} \|\mathcal{P}_u f\|_{L^p}^p \cdot \int_{\mathbb{S}^{d-1}} |\langle v, u\rangle|^p \, du \leq K_p \cdot \|f\|_{M^{p,\infty}}^p \cdot |v|^p, \tag{100}$$

proving (97).

□

3.2.4 Dilations

Suppose $\mathbb{B} \subset \mathbb{R}^d$ is the open unit ball in $\mathbb{R}^d$ centered at 0. Let $\Phi(w) = \alpha w$, where $\alpha > 1$. Then $\varepsilon(\Phi) = \alpha - 1$, and so bound (42) Theorem 3.1 is

$$\mathrm{SC}_{\eta,p}(f, f_\Phi) \leq 2^{(p-1)/p} \cdot \|f\|_{M_\eta^p} \cdot (\alpha - 1). \tag{101}$$

(We do not consider the bound (43), as for this choice of Φ it is never stronger than (42).)

We can prove a sharper estimate:

Theorem 3.11. *Let $1 \leq p \leq \infty$, and let f be in $L^p(\mathbb{B})$. Suppose $\alpha > 1$, and define f_α by $f_\alpha(w) = \alpha f(\alpha w)$. Then for any probability distribution η over $\mathbb{S}^{d-1}$,*

$$\mathrm{SC}_{\eta,p}(f, f_\alpha) \leq \|f\|_{M^p_\eta} \cdot \frac{\alpha - 1}{\alpha^{(p-1)/p}}. \tag{102}$$

The result follows from the following lemma:

Lemma 3.12. *Using the notation from the statement of Theorem 3.11,*

$$\mathrm{C}_p(\mathcal{P}_u f, \mathcal{P}_u f_\alpha) \leq \|\mathcal{P}_u(|f|)\|_{L^p} \cdot \frac{\alpha - 1}{\alpha^{(p-1)/p}}. \tag{103}$$

Theorem 3.4 follows immediately by taking the p-th power and averaging over all u.

Proof of Lemma 3.12. Without loss of generality, suppose $u = (1, 0, \ldots, 0)$. It is enough to show that for all $|t| < 1$,

$$\int \sup_{y \in \mathbb{R}^{d-1}} \chi(x, y, t)\, dx \leq \frac{\alpha - 1}{\alpha}; \tag{104}$$

this estimate can then be used in place of (56) in the proof of Lemma 3.3.

Without loss of generality, suppose $0 \leq t < 1$. Let S_t denote the set of all x, $|x| < 1$, satisfying $\sup_{y \in \mathbb{R}^{d-1}} \chi(x, y, t) = 1$. If $x \in S_t$, then $x < t < \alpha x$; since $t \geq 0$, this restricts $x \geq 0$ as well, and $S_t = (t/\alpha, t)$, so $|S_t| = t(1 - 1/\alpha)$, which is maximized at $t = 1$; thus

$$\int \sup_{y \in \mathbb{R}^{d-1}} \chi(x, y, t)\, dx \leq 1 - 1/\alpha = \frac{\alpha - 1}{\alpha}, \tag{105}$$

as claimed. Using this estimate in place of the bound $\int \sup_{y \in \mathbb{R}^{d-1}} \chi(x, y, t)\, dx \leq 2\varepsilon(\Phi)$ gives the final bound

$$\mathrm{C}(\mathcal{P}_u f, \mathcal{P}_u f_\Phi) \leq \|\mathcal{P}_u(|f|)\|_{L^p} (\alpha - 1)^{1/p} \left(\frac{\alpha - 1}{\alpha}\right)^{1/q} = \|\mathcal{P}_u(|f|)\|_{L^p} \cdot \frac{(\alpha - 1)}{\alpha^{(p-1)/p}}, \tag{106}$$

completing the proof.

□

3.3 Sliced Cramér distances and convolutions

In this section, we remark on the behavior of the sliced Cramér distance after convolution of its input functions. This situation occurs commonly in signal and image processing, where one typically observes signals that have been convolved with a function induced from the measurement apparatus. The bound we provide is analogous to Theorem 2.4 and Corollary 2.5 for Wasserstein and sliced Wasserstein distances, respectively. It essentially appears already in [74]; however, because this work states the result in terms of the distance between random variables (or equivalently, between probability distributions) we find it valuable to explicitly state and prove the generalization to arbitrary functions.

Theorem 3.13. *Let $1 \leq p \leq \infty$. Suppose $f, g, w : \mathbb{R}^d \to \mathbb{R}$ are compactly supported and in L^p, and let η be a probability measure over $\mathbb{S}^{d-1}$. Then*

$$\mathrm{SC}_{\eta,p}(f * w, g * w) \leq \|w\|_{M^{1,p}_\eta} \cdot \mathrm{SC}_{\eta,p}(f, g). \tag{107}$$

Remark 10. An immediate corollary to Theorem 3.13 is the bound $\mathrm{SC}_{\eta,p}(f * w, g * w) \leq \|w\|_{L^1} \cdot \mathrm{SC}_{\eta,p}(f, g)$. This essentially appears as Theorem 5 of [74] (though it is stated in terms of the distance between random variables). The special case where $d = 1$ is shown in [6], Theorem 2. Theorem 3.13 is a straightforward generalization of these results.

Proof of Theorem 3.13. We first show the result for $d = 1$, that is, for when f, g and w are functions on $\mathbb{R}$. We will show that

$$\mathrm{C}_p(f * w, g * w) \leq \|w\|_{L^1} \cdot \mathrm{C}_p(f, g). \tag{108}$$

Let H be in $\mathcal{A}_0$, with $\|H'\|_{L^q} = 1$. For any s, $\|H'(u+s)\|_{L^q(du)} = 1$ too. Then using Proposition 2.6,

$$\begin{aligned}
\langle f * w - g * w, H\rangle &= \int ((f-g) * w)(t) H(t)\, dt \\
&= \int \int (f-g)(t-s) w(s)\, ds\, H(t)\, dt \\
&= \int w(s) \int (f-g)(t-s) H(t)\, dt\, ds \\
&= \int w(s) \int (f-g)(u) H(u+s)\, du\, ds \\
&\le \|w\|_{L^1} \sup_s \left| \int (f-g)(u) H(u+s)\, du \right| \\
&\le \|w\|_{L^1} C_p(f,g),
\end{aligned} \tag{109}$$

and (108) now follows by taking the supremum over all such H and invoking Proposition 2.6.

We now turn to general $d \ge 1$. First, observe that $\mathcal{P}_u(w * h) = (\mathcal{P}_u w) * (\mathcal{P}_u h)$: indeed, by the Fourier slice theorem,

$$\begin{aligned}
(\mathcal{P}_u(w * h))\widehat{\ }(\xi) &= \widehat{(w * h)}(\xi u) \\
&= \widehat{w}(\xi u)\widehat{h}(\xi u) \\
&= \widehat{(\mathcal{P}_u w)}(\xi)\widehat{(\mathcal{P}_u h)}(\xi) \\
&= [(\mathcal{P}_u w) * (\mathcal{P}_u h)]\widehat{\ }(\xi),
\end{aligned} \tag{110}$$

and so, taking inverse Fourier transforms, $\mathcal{P}_u(w * h) = (\mathcal{P}_u w) * (\mathcal{P}_u h)$.

From the 1D bound, we then have

$$\begin{aligned}
\mathrm{C}_p(\mathcal{P}_u(w * f), \mathcal{P}_u(w * g)) &= \mathrm{C}_p((\mathcal{P}_u w) * (\mathcal{P}_u f), (\mathcal{P}_u w) * (\mathcal{P}_u g)) \\
&\le \|\mathcal{P}_u w\|_{L^1} \mathrm{C}_p(\mathcal{P}_u f, \mathcal{P}_u g).
\end{aligned} \tag{111}$$

Taking p-th powers and integrating over u gives the result.

□

4 Discretizations and robustness to noise

In this section, we will describe Fourier-based discretizations of the Cramér distance and the 2D sliced Cramér distance, with respect to the uniform measure over $\mathbb{S}^1$, between functions with equal integrals, and analyze their robustness to additive Gaussian noise. More precisely, we will show that, given vectors of noisy samples from two smooth functions, the discrete distance approximates the distances between the smooth functions only, removing the effect of the noise as the number of samples grows.

While they may be useful, these results are not especially surprising; indeed, the Cramér distance itself involves applying a smoothing filter to each input, which, by averaging the samples, naturally has a denoising effect. The denoising property is also of interest in contrasting with Wasserstein and sliced Wasserstein distances, which, because they are defined between probability measures, do not naturally induce distances between vectors sampled from a signal-plus-noise model.

We remark that in this section, we assume that the noiseless functions are C^∞. This assumption is made to simplify the analysis and statements of the theorems; the same results would hold under significantly weaker smoothness assumptions.

4.1 Robustness to noise in 1D

We define a discrete approximation to the 1D Volterra norm, which then yields an approximation to the Cramér distance. Let $a < b$ and let $L = b - a$ be the interval length. Let n a positive integer; we will assume for simplicity that n is even. Let $x \in \mathbb{R}^n$; the reader should think of x as having entries $x[j] = f(t_j)$, $j = 0, \dots, n-1$, where f is a function supported on $[a, b]$, and where $t_j = a + jL/n$.

Define the values $\alpha[k]$ (the normalized discrete Fourier coefficients of x) by

$$\alpha[k] = \frac{L}{n} \sum_{j=0}^{n-1} x[j] e^{-2\pi \mathrm{i} k t_j / L}, \tag{112}$$

for all integers k. Then for $-n/2 \leq k \leq n/2 - 1$, $\widehat{f}(k/L) \approx \alpha[k]$.

When $0 < |k| < n/2$, define

$$\beta[k] = \frac{\alpha[k]}{2\pi \mathrm{i} k / L}, \tag{113}$$

and when $k = 0$, define

$$\beta[0] = - \sum_{0<|k|<n/2} \beta[k] e^{2\pi \mathrm{i} k a / L}. \tag{114}$$

Then the $\beta[k]$ approximate the Fourier coefficients of $\mathcal{V}f$: $\widehat{(\mathcal{V}f)}(k/L) \approx \beta[k]$.

Define the function $\nu_x(t)$ by

$$\nu_x(t) = \frac{1}{L} \sum_{k=-n/2+1}^{n/2-1} \beta[k] e^{2\pi \mathrm{i} t k / L}. \tag{115}$$

(Note that we do not define $\beta[\pm n/2]$, because the terms they would contribute to $\nu_x(t)$ would either be purely imaginary or 0, depending on the convention.) Then for all t, $\nu_x(t) \approx (\mathcal{V}f)(t)$.

We then define the discrete Volterra p-norm of the vector x as follows:

$$\mathrm{V}_p(x) = \left(\frac{L}{n} \sum_{j=0}^{n-1} |\nu_x(t_j)|^p \right)^{1/p} \tag{116}$$

when $1 \leq p < \infty$, and

$$\mathrm{V}_\infty(x) = \max_{0 \leq j \leq n-1} |\nu_x(t_j)| \tag{117}$$

when $p = \infty$. Given two vectors x and y in $\mathbb{R}^n$, we then define their discrete Cramér distance as

$$\widehat{\mathrm{C}}_p(x, y) = \mathrm{V}_p(x - y). \tag{118}$$

Remark 11. Using the Fast Fourier Transform (FFT) and the inverse FFT (IFFT) to evaluate the $\alpha[k]$ and $\nu_x(t_j)$, respectively, the entire computation described here can be performed at cost $O(n \log n)$.

We can now state the main result from this section, which says that the discrete Cramér distance is robust to additive heteroscedastic Gaussian noise:

Theorem 4.1. *Suppose f and g are C^∞ functions on $\mathbb{R}$ that are supported on $[a, b]$, and satisfy $\int_a^b f = \int_a^b g$. Let $Z[0], Z[1], \ldots, Z[n-1]$, $\widetilde{Z}[0], \ldots, \widetilde{Z}[n-1]$ be independent Gaussians, where each $Z[j]$ has mean 0 and variance σ_j^2, and each $\widetilde{Z}[j]$ has mean 0 and variance $\widetilde{\sigma}_j^2$; and suppose σ satisfies*

$$\frac{1}{n} \sum_{j=0}^{n-1} \sigma_j^2 + \frac{1}{n} \sum_{j=0}^{n-1} \widetilde{\sigma}_j^2 \leq \sigma^2, \tag{119}$$

for all n. Let X_n and Y_n be vectors in $\mathbb{R}^n$ with entries $X_n[j] = f(t_j) + Z[j]$ and $Y_n[j] = g(t_j) + \widetilde{Z}[j]$. Then:

1. *Expected error: There are values $A, B > 0$ such that for all $\sigma \geq 0$, $n \geq 2$, and $1 \leq p \leq \infty$,*

$$\mathbb{E}\left[\left| \widehat{\mathrm{C}}_p(X_n, Y_n) - \mathrm{C}_p(f, g) \right| \right] \leq A\sigma \frac{\log(n)^{m(p)}}{\sqrt{n}} + \frac{B}{n}, \tag{120}$$

where $m(p) = \max\{1/2 - 1/p, 0\}$.

2. *Concentration bound. There are values $A, B, C > 0$ such that for all $\sigma > 0$, $t \geq 0$, and $n \geq C/t$,*

$$\mathbb{P}\left(\left|\widehat{\mathrm{C}}_p(X_n, Y_n) - \mathrm{C}_p(f,g)\right| \geq t\right) \leq A \cdot \exp\left(-B\frac{nt^2}{\sigma^2}\right) \tag{121}$$

for $1 \leq p \leq 2$, and

$$\mathbb{P}\left(\left|\widehat{\mathrm{C}}_p(X_n, Y_n) - \mathrm{C}_p(f,g)\right| \geq t\right) \leq A \cdot n \cdot \exp\left(-B\frac{nt^2}{\sigma^2}\right) \tag{122}$$

for $2 < p \leq \infty$.

3. *Almost sure limit. For all $1 \leq p \leq \infty$, $\widehat{\mathrm{C}}_p(X_n, Y_n) \to \mathrm{C}_p(f,g)$ almost surely as $n \to \infty$.*

Remark 12. It is straightforward to extend the definition of $\widehat{\mathrm{C}}_p(x,y)$, and the results from Theorem 4.1, to the setting where x and y contain samples of f and g taken on different grids, by interpolating the estimated Volterra transforms onto a common grid.

Remark 13. Because $\nu_x(t)$ can be evaluated at any t, not just the initial grid points, one could replace the definition of $\mathrm{V}_p(x)$ from (116) with any scheme for numerically integrating $|\nu_x(t)|^p$ over $[a,b]$.

Theorem 4.1 is an easy corollary of the following two results:

Proposition 4.2. *Suppose f is a C^∞ function on $\mathbb{R}$ that is supported on $[a,b]$ and satisfies $\int_a^b f = 0$. Let $x \in \mathbb{R}^n$ have entries $x[j] = f(t_j)$, $0 \leq j \leq n-1$. Then there is a value $C > 0$ such that*

$$|\mathrm{V}_p(x) - \|f\|_{V^p}| \leq \frac{C}{n} \tag{123}$$

for all $1 \leq p \leq \infty$ and $n \geq 2$.

Proposition 4.3. *Let $Z[0], Z[1], \dots, Z[n], \dots,$ be independent Gaussians, where each $Z[j]$ has mean 0 and variance σ_j^2; and suppose $\sigma > 0$ satisfies*

$$\frac{1}{n}\sum_{j=0}^{n-1} \sigma_j^2 \leq \sigma^2, \tag{124}$$

for all n. For each $n \geq 1$, let $Z_n = (Z[0], \dots, Z[n-1])$. Then:

1. *Expectation bound. There is a value $C > 0$ such that for all $\sigma \geq 0$, $n \geq 2$, and $1 \leq p \leq \infty$,*

$$\mathbb{E}\left[\mathrm{V}_p(Z_n)\right] \leq C\sigma\frac{\log(n)^{m(p)}}{\sqrt{n}}, \tag{125}$$

where $m(p) = \max\{1/2 - 1/p, 0\}$.

2. *Concentration bound. There are values $A, B > 0$ such that for all $\sigma > 0$, $t \geq 0$, and $n \geq 2$,*

$$\mathbb{P}\left\{\mathrm{V}_p(Z_n) \geq t\right\} \leq A \cdot \exp\left(-B\frac{nt^2}{\sigma^2}\right) \tag{126}$$

for $1 \leq p \leq 2$, and

$$\mathbb{P}\left\{\mathrm{V}_p(Z_n) \geq t\right\} \leq A \cdot n \cdot \exp\left(-B\frac{nt^2}{\sigma^2}\right) \tag{127}$$

for $2 < p \leq \infty$.

3. *Almost sure limit. For all $1 \leq p \leq \infty$, $\mathrm{V}_p(Z_n) \to 0$ almost surely as $n \to \infty$.*

We now show that Theorem 4.1 follows from Propositions 4.2 and 4.3:

Proof of Theorem 4.1. It is enough to show the analogous results for the Volterra norm of f alone; we can then replace f by $f-g$. Let $x[j]=f(t_j)$, $0\le j\le n-1$, so that $X_n=x+Z_n$. We have

$$\mathrm{V}_p(X_n)-\|f\|_{V_p}=\mathrm{V}_p(x+Z_n)-\|f\|_{V_p}\le \mathrm{V}_p(x)-\|f\|_{V_p}+\mathrm{V}_p(Z_n), \tag{128}$$

and

$$\mathrm{V}_p(X_n)-\|f\|_{V_p}=\mathrm{V}_p(x+Z_n)-\|f\|_{V_p}\ge \mathrm{V}_p(x)-\|f\|_{V_p}-\mathrm{V}_p(Z_n), \tag{129}$$

and therefore

$$\left|\mathrm{V}_p(X_n)-\|f\|_{V_p}\right|\le\left|\mathrm{V}_p(x)-\|f\|_{V_p}\right|+\mathrm{V}_p(Z_n). \tag{130}$$

The expected error bound (120) is then immediate, as is almost sure convergence. To show concentration, since $\left|\mathrm{V}_p(x)-\|f\|_{V_p}\right|\le C/n$ we have, for $1\le p\le 2$,

$$\begin{aligned}\mathbb{P}\{\left|\mathrm{V}_p(X_n)-\|f\|_{V_p}\right|\ge t\}&\le\mathbb{P}\{\left|\mathrm{V}_p(x)-\|f\|_{V_p}\right|+\mathrm{V}_p(Z_n)\ge t\}\\&\le\mathbb{P}\{\mathrm{V}_p(Z_n)\ge t-C/n\}\\&\le A\exp\left(-B\frac{n^2(t-C/n)^2}{\sigma^2}\right),\end{aligned} \tag{131}$$

and we then use $t-C/n\ge t/2$ when $n>2C/t$ to show (121). The case $2<p\le\infty$ is identical.

□

We now turn to the proofs of Propositions 4.2 and 4.3.

4.1.1 Proof of Proposition 4.2

Lemma 4.4. *Let $r>1$. Then there is a constant $C>0$ such that for all even $n\ge 2$ and $|k|<n/2$,*

$$\left|\alpha[k]-\widehat{f}(k/L)\right|\le\frac{C}{n^r}. \tag{132}$$

Proof. Write the Fourier series expansion of f:

$$f(t)=\frac{1}{L}\sum_{\ell=-\infty}^{\infty}\widehat{f}(\ell/L)e^{2\pi\mathrm{i}\ell t/L}. \tag{133}$$

For $|k|<n/2$, we have

$$\begin{aligned}\alpha[k]&=\frac{L}{n}\sum_{j=0}^{n-1}f(t_j)e^{-2\pi\mathrm{i}kt_j/L}\\&=\frac{1}{n}\sum_{j=0}^{n-1}\sum_{\ell=-\infty}^{\infty}\widehat{f}(\ell/L)e^{2\pi\mathrm{i}\ell t_j/L}e^{-2\pi\mathrm{i}kt_j/L}\\&=\frac{1}{n}\sum_{\ell=-\infty}^{\infty}\widehat{f}(\ell/L)\sum_{j=0}^{n-1}e^{2\pi\mathrm{i}(\ell-k)t_j/L}\\&=\frac{1}{n}\sum_{\ell=-\infty}^{\infty}\widehat{f}(\ell/L)e^{2\pi\mathrm{i}(\ell-k)a/L}\sum_{j=0}^{n-1}e^{2\pi\mathrm{i}(\ell-k)(t_j-a)/L}\\&=\frac{1}{n}\sum_{\ell=-\infty}^{\infty}\widehat{f}(\ell/L)e^{2\pi\mathrm{i}(\ell-k)a/L}\sum_{j=0}^{n-1}e^{2\pi\mathrm{i}(\ell-k)j/n}\\&=\sum_{\ell=-\infty}^{\infty}\widehat{f}(\ell/L)e^{2\pi\mathrm{i}(\ell-k)a/L}\delta_{\ell=k\bmod n}\\&=\sum_{\ell\,:\,\ell=k\bmod n}\widehat{f}(\ell/L)e^{2\pi\mathrm{i}(\ell-k)a/L}\\&=\widehat{f}(k/L)+\sum_{\ell:\ell\ne k,\,\ell=k\bmod n}\widehat{f}(\ell/L)e^{2\pi\mathrm{i}(\ell-k)a/L}.\end{aligned} \tag{134}$$

Because f is C^∞, $|\widehat{f}(k/L)| = O(|k|^{-r})$, and so

$$\begin{aligned}
\left|\alpha[k] - \widehat{f}(k/L)\right| &\le \sum_{\ell:\ell\neq k,\, \ell = k \bmod n} \left|\widehat{f}(\ell/L)\right| \\
&= \sum_{j\in\mathbb{Z}\setminus\{0\}} \left|\widehat{f}((k+jn)/L)\right| \\
&\le C \sum_{j\in\mathbb{Z}\setminus\{0\}} \frac{1}{|k+jn|^r} \\
&= \frac{C}{n^r} \sum_{j\in\mathbb{Z}\setminus\{0\}} \frac{1}{|k/n+j|^r} \\
&\le \frac{C}{n^r}, \qquad (135)
\end{aligned}$$

where the series is bounded by a constant because $|k/n| < 1/2$, and where C may change from line to line (but never depends on n). This completes the proof.

□

Corollary 4.5. *For all $r > 1$, there is a constant $C > 0$ such that for all $|t| < L$ and even $n \ge 2$,*

$$|\nu_x(t) - (\mathcal{V}f)(t)| \le \frac{C}{n^r}. \qquad (136)$$

Proof. Take any $q > r$. Applying Lemma 4.4 with q in place of r, we have:

$$\left|\sum_{0<|k|<n/2} \beta[k] e^{2\pi \mathrm{i} kt/L} - \sum_{0<|k|<n/2} \widehat{(\mathcal{V}f)}(k/L) e^{2\pi \mathrm{i} kt/L}\right| \le C \sum_{0<|k|<n/2} \frac{\left|\alpha[k] - \widehat{f}(k/L)\right|}{|k|} \le C\frac{\log(n)}{n^q}. \qquad (137)$$

Since f is C^∞, $|\widehat{f}(k/L)| = O(|k|^{-q})$, and therefore $|\widehat{(\mathcal{V}f)}(k/L)| = O(|k|^{-q-1})$, and so the tail may be bounded

$$\left|\sum_{|k|\ge n/2} \widehat{(\mathcal{V}f)}(k/L) e^{2\pi \mathrm{i} kt/L}\right| \le C \sum_{|k|\ge n/2} \frac{1}{|k|^{q+1}} \le \frac{C}{n^q}. \qquad (138)$$

Therefore, for any s,

$$\begin{aligned}
\left|\sum_{0<|k|<n/2} \beta[k] e^{2\pi \mathrm{i} ks/L} - \sum_{k\neq 0} \widehat{(\mathcal{V}f)}(k/L) e^{2\pi \mathrm{i} ks/L}\right| &\le \left|\sum_{0<|k|<n/2} (\beta[k] - \widehat{(\mathcal{V}f)}(k/L)) e^{2\pi \mathrm{i} ks/L}\right| + \left|\sum_{|k|\ge n/2} \widehat{(\mathcal{V}f)}(k/L) e^{2\pi \mathrm{i} ks/L}\right| \\
&\le \sum_{0<|k|<n/2} \left|\beta[k] - \widehat{(\mathcal{V}f)}(k/L))\right| + \sum_{|k|\ge n/2} \left|\widehat{(\mathcal{V}f)}(k/L)\right| \\
&\le C\frac{\log(n)}{n^q}. \qquad (139)
\end{aligned}$$

Taking $s = a$,

$$\left|\beta[0] - \widehat{(\mathcal{V}f)}(0)\right| = \left|\sum_{0<|k|<n/2} \beta[k] e^{2\pi \mathrm{i} ka/L} - \sum_{k\neq 0} \widehat{(\mathcal{V}f)}(k/L) e^{2\pi \mathrm{i} ka/L}\right| \le C\frac{\log(n)}{n^q}, \qquad (140)$$

and consequently, since $\log(n)/n^q \le C/n^r$,

$$|\nu_x(t) - (\mathcal{V}f)(t)| \le \left|\beta[0] - \widehat{(\mathcal{V}f)}(0)\right| + \left|\sum_{0<|k|<n/2} \beta[k] e^{2\pi \mathrm{i} ks/L} - \sum_{k\neq 0} \widehat{(\mathcal{V}f)}(k/L) e^{2\pi \mathrm{i} ks/L}\right| \le \frac{C}{n^r}. \qquad (141)$$

□

Lemma 4.6. *Suppose G has Lipschitz constant bounded by A on $[a,b]$, let $L = b-a$, and*

$$t_k = a + \frac{k}{n}L, \quad 0 \le k \le n. \tag{142}$$

Then for any $1 \le p < \infty$,

$$\left| \left(\frac{L}{n} \sum_{k=0}^{n-1} |G(t_k)|^p \right)^{1/p} - \|G\|_{L^p} \right| \le \frac{L^{1+1/p}}{n} A, \tag{143}$$

and, when $p = \infty$,

$$\left| \max_{0 \le k \le n} |G(t_k)| - \|G\|_{L^\infty} \right| \le \frac{L}{n} A. \tag{144}$$

Proof. First, suppose $1 \le p < \infty$. For brevity, let

$$T_n = \left(\frac{L}{n} \sum_{k=0}^{n-1} |G(t_k)|^p \right)^{1/p}. \tag{145}$$

For each $0 \le m \le n-1$, let

$$S_m = \left(\frac{L}{n} \right)^{1/p} |G(t_m)| \tag{146}$$

and

$$R_m = \left(\int_{t_m}^{t_{m+1}} |G(x)|^p \, dx \right)^{1/p}. \tag{147}$$

Then

$$T_n = \left(\sum_{m=0}^{n-1} |S_m|^p \right)^{1/p} \tag{148}$$

and

$$\|G\|_{L^p} = \left(\sum_{m=0}^{n-1} |R_m|^p \right)^{1/p}. \tag{149}$$

The Mean Value Theorem ensures that there is some t_m^* in the interval $[t_m, t_{m+1}]$ satisfying

$$R_m = \left(\frac{L}{n} \right)^{1/p} |G(t_m^*)|. \tag{150}$$

Since A bounds the Lipschitz constant for G,

$$\begin{aligned} |S_m - R_m| &= \left| S_m - \left(\frac{L}{n} \right)^{1/p} |G(t_m^*)| \right| \\ &= \left(\frac{L}{n} \right)^{1/p} \big| |G(t_m)| - |G(t_m^*)| \big| \\ &\le \left(\frac{L}{n} \right)^{1/p} A |t_m - t_m^*| \\ &\le A \left(\frac{L}{n} \right)^{1+1/p}. \end{aligned} \tag{151}$$

Consequently,

$$\begin{aligned}|T_n - \|G\|_{L^p}| &= \left| \left(\sum_{m=0}^{n-1} |S_m|^p \right)^{1/p} - \left(\sum_{m=0}^{n-1} |R_m|^p \right)^{1/p} \right| \\ &\le \left(\sum_{m=0}^{n-1} |S_m - R_m|^p \right)^{1/p} \\ &= \frac{L^{1+1/p}}{n} A. \end{aligned} \tag{152}$$

This completes the proof when p is finite. The proof for $p = \infty$ follows by taking the limit $p \to \infty$ and using the convergence of the p-norm to the ∞-norm.

□

We now complete the proof of Proposition 4.2. When $1 \le p < \infty$, from the triangle inequality

$$\begin{aligned}\left|\mathrm{V}_p(x) - \|f\|_{V^p}\right| &= \left| \left(\frac{L}{n} \sum_{j=0}^{n-1} |\nu_x(t_j)|^p \right)^{1/p} - \|f\|_{V^p} \right| \\ &\le \left| \left(\frac{L}{n} \sum_{j=0}^{n-1} |\nu_x(t_j)|^p \right)^{1/p} - \left(\frac{L}{n} \sum_{j=0}^{n-1} |(\mathcal{V}f)(t_j)|^p \right)^{1/p} \right| + \left| \left(\frac{L}{n} \sum_{j=0}^{n-1} |(\mathcal{V}f)(t_j)|^p \right)^{1/p} - \|f\|_{V^p} \right| \\ &\le \left(\frac{L}{n} \sum_{j=0}^{n-1} |\nu_x(t_j) - (\mathcal{V}f)(t_j)|^p \right)^{1/p} + \left| \left(\frac{L}{n} \sum_{j=0}^{n-1} |(\mathcal{V}f)(t_j)|^p \right)^{1/p} - \|\mathcal{V}f\|_{L^p} \right|. \end{aligned} \tag{153}$$

From Corollary 4.5, the first term is $O(1/n^r)$, $r > 1$; and from Lemma 4.6, the second term is $O(1/n)$. Since the constant is bounded with p, the case $p = \infty$ is obtained by taking the limit as $p \to \infty$.

4.1.2 Proof of Proposition 4.3

First, we note that the third part of the Proposition (almost sure convergence) follows immediately from the second part (the concentration bound) by using the following well-known corollary of the Borel-Cantelli Lemma (see e.g. Chapter 2, Section 10 of [63]):

Lemma 4.7. *Let $R_1, R_2, \ldots$ be a sequence of random numbers. Suppose that for all $\epsilon > 0$,*

$$\sum_{n=1}^{\infty} \mathbb{P}\{R_n > \epsilon\} < \infty. \tag{154}$$

Then $R_n \to 0$ almost surely.

Recall that the vectors α and β are defined as follows:

$$\alpha[k] = \frac{L}{n} \sum_{j=0}^{n-1} Z[j] e^{-2\pi \mathrm{i} k t_j / L}, \quad -n/2 \le k \le n/2 - 1, \tag{155}$$

$$\beta[k] = \frac{\alpha[k]}{2\pi \mathrm{i} k / L}, \quad 0 < |k| < n/2, \tag{156}$$

and when $k = 0$,

$$\beta[0] = - \sum_{0<|k|<n/2} \beta[k] e^{2\pi \mathrm{i} k a / L}. \tag{157}$$

Define the random vector W by $W[j] = \nu_Z(t_j)$, that is,

$$W[j] = \frac{1}{L} \sum_{k=-n/2+1}^{n/2-1} \beta[k] e^{2\pi \mathrm{i} t_j k/L}, \tag{158}$$

for $0 \le j \le n-1$. Then for $1 \le p < \infty$,

$$\mathrm{V}_p(Z_n) = \left(\frac{L}{n} \sum_{j=0}^{n-1} |W[j]|^p \right)^{1/p}, \tag{159}$$

and $\mathrm{V}_\infty(Z_n) = \|W\|_\infty$.

It will be convenient to define the auxiliary vector X by

$$X[j] = \sum_{0<|k|<n/2} \beta[k] e^{2\pi \mathrm{i} t_j k/L}. \tag{160}$$

Then $W[j] = (X[j] + \beta[0])/L$.

Expectation of $\mathrm{V}_p(Z_n)$**,** $1 \le p \le 2$**.** The following result may be found (in more general form) in Chapter V of [80]:

Lemma 4.8. *There is a constant* $C > 0$ *such that for any positive integer* m *and real number* A*,*

$$\left| \sum_{k=1}^{m} \frac{\sin(kA)}{k} \right| \le C. \tag{161}$$

We use Lemma 4.8 to bound the variance of each entry of X. For a fixed $0 \le \ell \le n-1$,

$$\begin{aligned} X[\ell] &= \sum_{0<|k|<n/2} \beta[k] e^{2\pi \mathrm{i} t_\ell k/L} \\ &= \frac{L}{2\pi \mathrm{i}} \sum_{0<|k|<n/2} \frac{\alpha[k]}{k} e^{2\pi \mathrm{i} t_\ell k/L} \\ &= \frac{L}{2\pi \mathrm{i}} \sum_{0<|k|<n/2} \frac{1}{k} \left(\frac{L}{n} \sum_{j=0}^{n-1} Z[j] e^{-2\pi \mathrm{i} k t_j/L} \right) e^{2\pi \mathrm{i} t_\ell k/L} \\ &= \frac{L^2}{2\pi \mathrm{i} n} \sum_{j=0}^{n-1} Z[j] \sum_{0<|k|<n/2} \frac{e^{2\pi \mathrm{i} k(\ell-j)/n}}{k} \\ &= \frac{L^2}{2\pi \mathrm{i} n} \sum_{j=0}^{n-1} Z[j] \sum_{k=1}^{n/2-1} \frac{\sin(2\pi(\ell-j)k/n)}{k}. \end{aligned} \tag{162}$$

By Lemma 4.8, and the independence of the $Z[j]$,

$$\mathbb{E}\left[|X[\ell]|^2 \right] = C \frac{L^4}{n^2} \sum_{j=0}^{n-1} \sigma_j^2 \left(\sum_{k=1}^{n/2-1} \frac{\sin(2\pi(\ell-j)k/n)}{k} \right)^2 \le C L^4 \frac{\sigma^2}{n}. \tag{163}$$

Similarly, we have

$$\beta[0] = -\frac{L}{2\pi \mathrm{i}} \sum_{0<|k|<n/2} \frac{\alpha[k]}{k} e^{2\pi \mathrm{i} k a/L} = -\frac{L^2}{2\pi \mathrm{i} n} \sum_{j=0}^{n-1} Z[j] \sum_{k=1}^{n/2-1} \frac{\sin(2\pi jk/n)}{k}, \tag{164}$$

and again using Lemma 4.8,

$$\mathbb{E}\left[|\beta[0]|^2\right] = C\frac{L^4}{n^2}\sum_{j=0}^{n-1}\sigma_j^2\left(\sum_{k=1}^{n/2-1}\frac{\sin(2\pi jk/n)}{k}\right)^2 \leq CL^4\frac{\sigma^2}{n}. \tag{165}$$

Since $W[j] = (X[j] + \beta[0])/L$, it follows that

$$\mathbb{E}[W[j]^2] \leq CL^2\frac{\sigma^2}{n}. \tag{166}$$

Consequently,

$$\mathbb{E}[\mathrm{V}_2(Z_n)^2] = \frac{L}{n}\sum_{j=0}^{n-1}\mathbb{E}[W[j]^2] \leq CL^3\frac{\sigma^2}{n}, \tag{167}$$

and therefore, using Jensen's inequality,

$$\mathbb{E}[\mathrm{V}_2(Z_n)] \leq \sqrt{\mathbb{E}[\mathrm{V}_2(Z_n)^2]} \leq CL^{3/2}\frac{\sigma}{\sqrt{n}}. \tag{168}$$

Since $\mathrm{V}_p(Z_n) \leq L^{1/p-1/2}\mathrm{V}_2(Z_n)$ for $1 \leq p \leq 2$, we then have

$$\mathbb{E}[\mathrm{V}_p(Z_n)] \leq CL^{1+1/p}\frac{\sigma}{\sqrt{n}}. \tag{169}$$

Expectation of $\mathrm{V}_p(Z_n)$, $2 < p \leq \infty$. We start with a simple bound on the moment generating function of the absolute value of a Gaussian random variable, whose proof is provided for the reader's convenience:

Lemma 4.9. *Suppose Y is a mean zero mean Gaussian with variance τ. Then $\mathbb{E}\left[e^{s|Y|}\right] \leq 2e^{\tau s^2/2}$ for all s.*

Proof. Since the moment generating function of Y itself is $\mathbb{E}[e^{sY}] = e^{\tau s^2/2}$, we have the bound

$$\begin{aligned}\mathbb{E}\left[e^{s|Y|}\right] &= \frac{1}{\sqrt{2\pi\tau}}\int_{-\infty}^{\infty} e^{-y^2/2\tau}e^{s|y|}\,dy \\ &= 2\frac{1}{\sqrt{2\pi\tau}}\int_{0}^{\infty} e^{-y^2/2\tau}e^{sy}\,dy \\ &\leq 2\frac{1}{\sqrt{2\pi\tau}}\int_{-\infty}^{\infty} e^{-y^2/2\tau}e^{sy}\,dy \\ &= 2e^{\tau s^2/2}. \end{aligned} \tag{170}$$

□

From (163), the variance of $X[\ell]$ is bounded above by $CL^4\sigma^2/n$. Therefore, for any $s > 0$, using Jensen's inequality, we have

$$\begin{aligned} e^{\mathbb{E}[s\|X\|_\infty]} &\leq \mathbb{E}\left[e^{s\|X\|_\infty}\right] \\ &= \mathbb{E}\left[\max_{\ell} e^{s|X[\ell]|}\right] \\ &\leq \mathbb{E}\left[\sum_{\ell=0}^{n-1} e^{s|X[\ell]|}\right] \\ &= \sum_{\ell=0}^{n-1}\mathbb{E}\left[e^{s|X[\ell]|}\right] \\ &\leq n\max_{0\leq\ell\leq n-1}\mathbb{E}\left[e^{s|X[\ell]|}\right] \\ &\leq 2n\max_{0\leq\ell\leq n-1} e^{s^2\mathrm{Var}(X[\ell])/2} \\ &\leq 2ne^{Cs^2L^4\sigma^2/n} \end{aligned} \tag{171}$$

and so, taking the logarithm of each side,

$$\mathbb{E}\left[\|X\|_\infty\right] \leq C\left(\frac{\log(n)}{s} + \frac{sL^4\sigma^2}{n}\right), \tag{172}$$

and taking $s = \sigma^{-1}L^{-2}\sqrt{n\log(n)}$ then gives the bound

$$\mathbb{E}\left[\|X\|_\infty\right] \leq CL^2\sigma\sqrt{\frac{\log(n)}{n}}. \tag{173}$$

It is straightforward to show the same bound for $\beta[0]$, and since $W[\ell] = (X[\ell] + \beta[0])/L$ and $\mathrm{V}_\infty(Z_n) = \|W\|_\infty$, it follows that

$$\mathbb{E}\left[\mathrm{V}_\infty(Z_n)\right] \leq CL\sigma\sqrt{\frac{\log(n)}{n}}. \tag{174}$$

Now, for any $2 < p < \infty$,

$$\mathrm{V}_p(Z_n) = \left(\frac{L}{n}\sum_{j=0}^{n-1}|W[j]|^p\right)^{1/p} \leq \|W\|_\infty^{1-2/p}\left(\frac{L}{n}\sum_{j=0}^{n-1}|W[j]|^2\right)^{1/p} = \mathrm{V}_2(Z_n)^{2/p}\cdot\mathrm{V}_\infty(Z_n)^{1-2/p}, \tag{175}$$

and so with $r = p/2$ and $s = (1-2/p)^{-1}$, by Hölder's inequality

$$\begin{aligned}
\mathbb{E}\left[\mathrm{V}_p(Z_n)\right] &\leq \mathbb{E}\left[\mathrm{V}_2(Z_n)^{2/p}\cdot\mathrm{V}_\infty(Z_n)^{1-2/p}\right] \\
&\leq \left(\mathbb{E}\left[\mathrm{V}_2(Z_n)^{2r/p}\right]\right)^{1/r}\cdot\left(\mathbb{E}\left[\mathrm{V}_\infty(Z_n)^{s(1-2/p)}\right]\right)^{1/s} \\
&= \left(\mathbb{E}\left[\mathrm{V}_2(Z_n)\right]\right)^{2/p}\cdot\left(\mathbb{E}\left[\mathrm{V}_\infty(Z_n)\right]\right)^{1-2/p} \\
&= CL^{1+1/p}\frac{\sigma\log(n)^{1/2-1/p}}{\sqrt{n}}.
\end{aligned} \tag{176}$$

Concentration of $\mathrm{V}_p(Z_n)$, $1 \leq p \leq 2$. First, observe that for each $k \neq 0$,

$$\begin{aligned}
|\alpha[k]|^2 &= \frac{L^2}{n^2}\sum_{i,j}Z[i]Z[j]e^{-2\pi\mathrm{i}kt_i/L}e^{2\pi\mathrm{i}kt_j/L} \\
&= \frac{L^2}{n^2}\sum_{i=0}^{n-1}Z[i]^2 + \frac{L^2}{n^2}\sum_{i\neq j}Z[i]Z[j]e^{-2\pi\mathrm{i}kt_i/L}e^{2\pi\mathrm{i}kt_j/L},
\end{aligned} \tag{177}$$

and taking expectations then gives

$$\mathbb{E}\left[|\alpha[k]|^2\right] = \frac{L^2}{n^2}\sum_{i=0}^{n-1}\sigma_j^2 \leq L^2\frac{\sigma^2}{n}. \tag{178}$$

Therefore, for $k \neq 0$,

$$\mathbb{E}\left[|\beta[k]|^2\right] \leq CL^4\frac{\sigma^2}{nk^2}. \tag{179}$$

Let $\varphi[k] = \mathrm{Re}(\beta[k])$ and $\psi[k] = \mathrm{Im}(\beta[k])$, $0 < k < n/2$. Also let $c_k[j] = \cos(2\pi kt_j/L)$ and $s_k[j] = \sin(2\pi kt_j/L)$, $1 \leq k \leq n/2-1$. All $n-2$ vectors c_k and s_k are pairwise orthogonal with $\|c_k\|_2 = \|s_k\|_2 = \sqrt{n/2}$. Then

$$X[j] = 2\sum_{0<k<n/2}\left[\varphi[k]c_k[j] - \psi[k]s_k[j]\right], \tag{180}$$

and so $\|X\|_2^2 = 2n(\|\varphi\|_2^2 + \|\psi\|_2^2)$, and therefore

$$\begin{aligned} \mathrm{V}_2(Z_n)^2 &= (L/n)\|W\|_2^2 \\ &= (L/n)\|X/L\|_2^2 + (L/n)|\beta[0]/L|^2 \\ &= \|X\|_2^2/(nL) + |\beta[0]|^2/(nL) \\ &= (2/L)\|\varphi\|_2^2 + (2/L)\|\psi\|_2^2 + |\beta[0]|^2/(nL). \end{aligned} \tag{181}$$

The $n-2$ variables $\varphi[k]$ and $\psi[k]$ are independent Gaussians with zero mean, and respective variances

$$\tau_\varphi[k] \equiv \mathbb{E}\left[\varphi[k]^2\right] \le CL^4 \frac{\sigma^2}{nk^2} \tag{182}$$

and

$$\tau_\psi[k] \equiv \mathbb{E}\left[\psi[k]^2\right] \le CL^4 \frac{\sigma^2}{nk^2}, \tag{183}$$

where the bounds follow from (179). Let $\tau = (\tau_\varphi, \tau_\psi)$. Then

$$\|\tau\|_\infty \le CL^4 \frac{\sigma^2}{n} \tag{184}$$

and

$$\|\tau\|_2^2 = \|\tau_\varphi\|_2^2 + \|\tau_\psi\|_2^2 \le CL^8 \frac{\sigma^4}{n^2} \sum_{k=1}^{n/2-1} \frac{1}{k^4} \le CL^8 \frac{\sigma^4}{n^2}. \tag{185}$$

Furthermore,

$$\mathbb{E}\left[\|\varphi\|_2^2 + \|\psi\|_2^2\right] = \sum_{k=0}^{n/2-1} \tau_\varphi[k] + \sum_{k=0}^{n/2-1} \tau_\psi[k] \le \sum_{k=0}^{n/2-1} CL^4 \frac{\sigma^2}{nk^2} \le CL^4 \frac{\sigma^2}{n}. \tag{186}$$

Let $\Delta = \|\varphi\|_2^2 + \|\psi\|_2^2$. From Lemma 1 in [33], for any $t > 0$

$$\mathbb{P}\left\{\Delta - \mathbb{E}[\Delta] \ge 2\|\tau\|_2\sqrt{t} + 2\|\tau\|_\infty t\right\} \le \exp(-t), \tag{187}$$

which easily implies the one-sided subexponential tail bound, namely that for any $s > 0$,

$$\mathbb{P}\left\{\Delta - \mathbb{E}[\Delta] \ge s\right\} \le \exp\left(-C \min\left\{\frac{s^2}{\|\tau\|_2^2}, \frac{s}{\|\tau\|_\infty}\right\}\right) \le \exp\left(-C \min\left\{\frac{n^2 s^2}{L^8\sigma^4}, \frac{ns}{L^4\sigma^2}\right\}\right), \tag{188}$$

where C is a universal constant. Since $\mathbb{E}[\Delta] \le CL^4\sigma^2/n$, it then follows that for any $s > 0$,

$$\mathbb{P}\left\{(2/L)\|\varphi\|_2^2 + (2/L)\|\psi\|_2^2 \ge s\right\} \le \exp\left(-C \min\left\{\frac{n^2 s^2}{L^6\sigma^4}, \frac{ns}{L^3\sigma^2}\right\}\right) \tag{189}$$

for all $n \ge L^3\sigma^2/s$. It is straightforward to prove the same bound for $(1/nL)\beta[0]^2$; and so by (181),

$$\mathbb{P}\left\{\mathrm{V}_2(Z_n)^2 \ge s\right\} \le 2\exp\left(-C \min\left\{\frac{n^2 s^2}{L^6\sigma^4}, \frac{ns}{L^3\sigma^2}\right\}\right), \tag{190}$$

and consequently, for all $t > 0$

$$\mathbb{P}\left\{\mathrm{V}_2(Z_n) \ge t\right\} \le 2\exp\left(-C \min\left\{\frac{n^2 t^4}{L^6\sigma^4}, \frac{nt^2}{L^3\sigma^2}\right\}\right) \le 2\exp\left(-C\frac{nt^2}{L^3\sigma^2}\right) \tag{191}$$

for all $n \ge L^3\sigma^2/t^2$. When $n < L^3\sigma^2/t^2$, then the right side of (191) is bounded below by a postive constant, and so for sufficiently large $A > 0$ the bound

$$\mathbb{P}\left\{\mathrm{V}_p(Z_n) \ge t\right\} \le A\exp\left(-C\frac{nt^2}{L^3\sigma^2}\right) \tag{192}$$

is valid for all $t \ge 0$ and $n \ge 2$.

Since $\mathrm{V}_p(Z_n) \le L^{1/p-1/2}\mathrm{V}_2(Z_n)$ for all $1 \le p \le 2$, it follows that

$$\mathbb{P}\left\{\mathrm{V}_p(Z_n) \ge t\right\} \le A\exp\left(-C\frac{nt^2}{L^{2+2/p}\sigma^2}\right) \tag{193}$$

for all $t \ge 0$ and $n \ge 2$.

Concentration of $\mathrm{V}_p(Z_n)$**,** $2 < p \leq \infty$**.** Since each $W[j]$ is Gaussian with variance bounded by $CL^2\sigma^2/n$ (by (166)), standard Gaussian tails bounds imply

$$\mathbb{P}\left(|W[j]| \geq t\right) \leq 2\exp\left(-C\frac{nt^2}{L^2\sigma^2}\right) \tag{194}$$

(e.g. see Chapter 2 in [72]). Since $\mathrm{V}_\infty(Z_n) = \|W\|_\infty$, by the union bound, therefore,

$$\mathbb{P}\left(\mathrm{V}_\infty(Z_n) \geq t\right) \leq n \max_{0\leq j\leq n-1} \mathbb{P}\left(|W[j]| \geq t\right) \leq 2n\exp\left(-C\frac{nt^2}{L^2\sigma^2}\right). \tag{195}$$

Furthermore, for all $2 < p < \infty$, since $\mathrm{V}_p(Z_n) \leq L^{1/p}\mathrm{V}_\infty(Z_n)$ we have the bound

$$\mathbb{P}\left(\mathrm{V}_p(Z_n) \geq t\right) \leq 2n\exp\left(-C\frac{nt^2}{L^{2+2/p}\sigma^2}\right), \tag{196}$$

as desired.

4.2 Robustness to noise in 2D

We turn now to the discrete approximation of the sliced Cramér distance in 2D, with respect to the uniform measure over $\mathbb{S}^1$. We follow the discretization described in [61], which uses a non-uniform discrete Fourier transform to compute the Radon transform of the input functions. Let $R > 0$ and $L = 2R$. Let n a positive integer; throughout this discussion, we will assume for simplicity that n is even. Let $x \in \mathbb{R}^{n^2} = \mathbb{R}^n \times \mathbb{R}^n$; the reader should think of x as having entries $x[i,j] = f(t_i, t_j)$, $i, j = 0, \ldots, n-1$, where $f : \mathbb{R}^2 \to \mathbb{R}$ is a function supported on the disc of radius R centered at the origin and where $t_j = -R + 2Rj/n$.

For $\theta \in [0, \pi)$, define the values

$$\alpha_\theta[k] = \frac{L^2}{n^2}\sum_{i,j} x[i,j]e^{-2\pi \mathrm{i}k(t_i\cos(\theta)+t_j\sin(\theta))/L}. \tag{197}$$

Then for $-n/2 \leq k \leq n/2$, $\alpha_\theta[k] \approx \widehat{f}((k/L)\cos(\theta), (k/L)\sin(\theta)) = \widehat{(\mathcal{P}_\theta f)}(k/L)$.

For $0 < |k| < n/2$, let

$$\beta_\theta[k] = \frac{\alpha_\theta[k]}{2\pi \mathrm{i}k/L}. \tag{198}$$

For $k = 0$, define

$$\beta_\theta[0] = -\sum_{0<|k|<n/2} \beta_\theta[k]e^{-2\pi \mathrm{i}kR/L} = -\sum_{0<|k|<n/2} \beta_\theta[k](-1)^k. \tag{199}$$

Define $\nu_x(t,\theta)$ by

$$\nu_x(t,\theta) = \frac{1}{L}\sum_{k=-n/2+1}^{n/2-1} \beta_\theta[k]e^{2\pi \mathrm{i}tk/L}. \tag{200}$$

Let $\theta_\ell = \pi\ell/n$, $\ell = 0, \ldots, n-1$. We then define the estimated sliced Volterra norm for $1 \leq p < \infty$ to be

$$\mathrm{SV}_p(x) = \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\frac{L}{n}\sum_{j=0}^{n}|\nu_x(t_j,\theta_\ell)|^p\right)^{1/p}, \tag{201}$$

and

$$\mathrm{SV}_\infty(x) = \max_{0\leq j,\ell\leq n-1}|\nu_x(t_j,\theta_\ell)|. \tag{202}$$

Given two vectors x and y, define their sliced Cramér distance to be

$$\widehat{\mathrm{SC}}_p(x,y) = \mathrm{SV}_p(x-y). \tag{203}$$

Remark 14. Using a non-uniform Fast Fourier Transform (NUFFT) (for example, see [16, 17, 18, 24, 4, 5]) to evaluate the $\alpha_{\theta_\ell}[k]$ and the $\nu_x(t_j, \theta_\ell)$, the entire computation described here can be performed at cost $O(n^2 \log n)$. In our implementation, we use the Flatiron Institute NUFFT (FINUFFT) [4, 5].

We can now state the main result from this section, which says that the discrete sliced Cramér distance in 2D is robust to additive heteroscedastic Gaussian noise:

Theorem 4.10. *Suppose f and g are C^∞ functions on $\mathbb{R}^2$ that are supported on the disc of radius $R > 0$ centered at the origin and satisfy $\int_{\mathbb{R}^2} f = \int_{\mathbb{R}^2} g$. Let $Z[j,k]$, $\widetilde{Z}[j,k]$, $0 \le j, k \le n-1$, be independent Gaussians, where each $Z[j,k]$ has mean 0 and variance σ_{jk}^2, and each $\widetilde{Z}[j,k]$ has mean 0 and variance $\widetilde{\sigma}_{jk}^2$; and suppose that for all n, $\sigma > 0$ satisfies*

$$\frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} \sigma_{jk}^2 + \frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} \widetilde{\sigma}_{jk}^2 \le \sigma^2. \tag{204}$$

Let X_n and Y_n be vectors in $\mathbb{R}^{n^2}$ with entries $X_n[j,k] = f(t_j, t_k) + Z[j,k]$ and $Y_n[j,k] = g(t_j, t_k) + \widetilde{Z}[j,k]$. Then:

1. *Expected error: There are values $A, B > 0$ such that for all $\sigma \ge 0$, $n \ge 2$, and $1 \le p \le \infty$,*

$$\mathbb{E}\left[\left|\widehat{\mathrm{SC}}_p(X_n, Y_n) - \mathrm{SC}_p(f,g)\right|\right] \le A\sigma \frac{\log(n)^{m(p)}}{n} + \frac{B}{n}, \tag{205}$$

   where $m(p) = \max\{1/2 - 1/p, 0\}$.

2. *Concentration bound. There are values $A, B, C > 0$ such that for all $\sigma > 0$, $t \ge 0$, and $n \ge C/t$,*

$$\mathbb{P}\left(\left|\widehat{\mathrm{SC}}_p(X_n, Y_n) - \mathrm{SC}_p(f,g)\right| \ge t\right) \le A \cdot \exp\left(-B\frac{n^2 t^2}{\sigma^2}\right) \tag{206}$$

   for $1 \le p \le 2$, and

$$\mathbb{P}\left(\left|\widehat{\mathrm{SC}}_p(X_n, Y_n) - \mathrm{SC}_p(f,g)\right| \ge t\right) \le A \cdot n^2 \cdot \exp\left(-B\frac{n^2 t^2}{\sigma^2}\right) \tag{207}$$

   for $2 < p \le \infty$.

3. *Almost sure limit. For all $1 \le p \le \infty$, $\widehat{\mathrm{SC}}_p(X_n, Y_n) \to \mathrm{SC}_p(f,g)$ almost surely as $n \to \infty$.*

As in the 1D case, Theorem 4.10 is an easy corollary of the following two results:

Proposition 4.11. *Suppose f is a C^∞ function on $\mathbb{R}^2$ that is supported on the disc of radius $R > 0$ centered at the origin, and $\int_{\mathbb{R}^2} f = 0$. Let $1 \le p \le \infty$, and let $x \in \mathbb{R}^n$ have entries $x[j,k] = f(t_j, t_k)$, $0 \le j, k \le n-1$. Then*

$$|\mathrm{SV}_p(x) - \|f\|_{SV^p}| \le \frac{C}{n}, \tag{208}$$

where $C > 0$ does not depend on n.

Proposition 4.12. *Let $Z[j,k]$, $j, k \ge 0$, be independent Gaussians, where each $Z[j,k]$ has mean 0 and variance σ_{jk}^2; and suppose $\sigma > 0$ satisfies*

$$\frac{1}{n^2} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \sigma_{jk}^2 \le \sigma^2, \tag{209}$$

for all n. For each $n \ge 1$, let $Z_n \in \mathbb{R}^{n^2}$ have entries $Z[j,k]$. Then:

1. *Expectation bound. There is a value $C > 0$ such that for all $\sigma \ge 0$, $n \ge 2$, and $1 \le p \le \infty$,*

$$\mathbb{E}\left[\mathrm{SV}_p(Z_n)\right] \le C\sigma \frac{\log(n)^{m(p)}}{n}, \tag{210}$$

   where $m(p) = \max\{1/2 - 1/p, 0\}$.

2. Concentration bound. There are values $A, B > 0$ such that for all $\sigma > 0$, $t \geq 0$, and $n \geq 2$,

$$\mathbb{P}\left(\mathrm{SV}_p(Z_n) \geq t\right) \leq A \cdot \exp\left(-B\frac{n^2t^2}{\sigma^2}\right) \tag{211}$$

for all $1 \leq p \leq 2$, and

$$\mathbb{P}\left(\mathrm{SV}_p(Z_n) \geq t\right) \leq A \cdot n^2 \cdot \exp\left(-B\frac{n^2t^2}{\sigma^2}\right) \tag{212}$$

for all $2 < p \leq \infty$.

3. Almost sure limit. For all $1 \leq p \leq \infty$, $\mathrm{SV}_p(Z_n) \to 0$ almost surely as $n \to \infty$.

4.2.1 Proof of Proposition 4.11

Lemma 4.13. *For all $r > 2$, there is a $C > 0$ such that*

$$\left|\alpha_\theta[m] - \widehat{(\mathcal{P}_\theta f)}(m/L)\right| \leq \frac{C}{n^r} \tag{213}$$

for all $0 \leq \theta < \pi$, $n \geq 2$, and $|m| < n/2$.

Proof. We write the Fourier expansion of f:

$$f(s,t) = \frac{1}{L^2}\sum_{k,\ell} \widehat{f}(k/L, \ell/L)e^{2\pi\mathrm{i}(sk+t\ell)/L}. \tag{214}$$

Suppose $|\xi| \leq n/(2L)$ and $|\nu| \leq n/(2L)$. Let $f_{\xi,\nu}(s,t) = f(s,t)e^{-2\pi\mathrm{i}(s\xi+t\nu)}$. Then $\widehat{f_{\xi,\nu}}(\varphi,\psi) = \widehat{f}(\varphi+\xi,\psi+\nu)$, and so

$$\begin{aligned} f(s,t)e^{-2\pi\mathrm{i}(s\xi+t\nu)} &= f_{\xi,\nu}(s,t) \\ &= \frac{1}{L^2}\sum_{k,\ell} \widehat{f_{\xi,\nu}}(k/L,\ell/L)e^{2\pi\mathrm{i}(sk+t\ell)/L} \\ &= \frac{1}{L^2}\sum_{k,\ell} \widehat{f}(k/L+\xi,\ell/L+\nu)e^{2\pi\mathrm{i}(sk+t\ell)/L}, \end{aligned} \tag{215}$$

and consequently

$$f(s,t) = \frac{1}{L^2}\sum_{k,\ell} \widehat{f}(k/L+\xi,\ell/L+\nu)e^{2\pi\mathrm{i}(s(k/L+\xi)+t(\ell/L+\nu))}. \tag{216}$$

Therefore,

$$
\begin{aligned}
\frac{L^2}{n^2}\sum_{i,j} f(t_i,t_j)e^{-2\pi \mathrm{i}(t_i\xi+t_j\nu)} &= \frac{1}{n^2}\sum_{i,j}\sum_{k,\ell}\widehat{f}(k/L+\xi,\ell/L+\nu)e^{2\pi\mathrm{i}(t_i(k/L+\xi)+t_j(\ell/L+\nu))}e^{-2\pi\mathrm{i}(t_i\xi+t_j\nu)}\\
&= \frac{1}{n^2}\sum_{i,j}\sum_{k,\ell}\widehat{f}(k/L+\xi,\ell/L+\nu)e^{2\pi\mathrm{i}(t_ik/L+t_j\ell/L)}\\
&= \sum_{k,\ell}\frac{1}{n^2}\sum_{i,j}\widehat{f}(k/L+\xi,\ell/L+\nu)e^{2\pi\mathrm{i}((Li/n-R)k/L+(Lj/n-R)\ell/L)}\\
&= \sum_{k,\ell}\frac{1}{n^2}\sum_{i,j}\widehat{f}(k/L+\xi,\ell/L+\nu)e^{2\pi\mathrm{i}(-Rk/L+ki/n-R\ell/L+\ell j/n)}\\
&= \sum_{k,\ell}\widehat{f}(k/L+\xi,\ell/L+\nu)(-1)^{k+\ell}\frac{1}{n^2}\sum_{i,j}e^{2\pi\mathrm{i}(ki+\ell j)/n}\\
&= \sum_{k,\ell}\widehat{f}(k/L+\xi,\ell/L+\nu)(-1)^{k+\ell}\delta_{k=0 \bmod n}\delta_{\ell=0 \bmod n}\\
&= \sum_{c,d}\widehat{f}(cn/L+\xi,dn/L+\nu)(-1)^{(c+d)/n}\\
&= \widehat{f}(\xi,\nu)+\sum_{(c,d)\in\mathbb{Z}^2\setminus\{(0,0)\}}\widehat{f}(cn/L+\xi,dn/L+\nu)(-1)^{(c+d)/n}. \qquad (217)
\end{aligned}
$$

Since f is C^∞, $|\widehat{f}(\xi)| = O(|\xi|^{-r})$, and so the series can be bounded

$$
\begin{aligned}
\sum_{(c,d)\in\mathbb{Z}^2\setminus\{(0,0)\}}\left|\widehat{f}(cn/L+\xi,dn/L+\nu)\right| &\le C\sum_{(c,d)\in\mathbb{Z}^2\setminus\{(0,0)\}}\left(\frac{1}{(cn/L+\xi)^2+(dn/L+\nu)^2}\right)^{r/2}\\
&= C\frac{L^r}{n^r}\sum_{(c,d)\in\mathbb{Z}^2\setminus\{(0,0)\}}\left(\frac{1}{(c+\xi L/n)^2+(d+\nu L/n)^2}\right)^{r/2}. \qquad (218)
\end{aligned}
$$

Since $|\xi| \le n/(2L)$ and $|\nu| \le n/(2L)$ and $r > 2$, the series is bounded by a constant. Taking $\xi = (m/L)\cos(\theta)$ and $\nu = (m/L)\sin(\theta)$, and using that $m \le n/2$, completes the proof.

□

Corollary 4.14. *For all $r > 2$, there is $C > 0$ such that*

$$
|\nu_x(t,\theta) - (\mathcal{V}\mathcal{P}_\theta f)(t)| \le \frac{C}{n^r} \qquad (219)
$$

for all $0 \le \theta < \pi$ and $|t| < L$.

Proof. Let $q > r$. For all t and θ, applying Lemma 4.13 with q in place of r gives

$$
\begin{aligned}
\left|\frac{1}{L}\sum_{0<|k|\le n/2-1}\beta_\theta[k]e^{2\pi\mathrm{i}tk/L} - \frac{1}{L}\sum_{0<|k|\le n/2-1}\frac{\widehat{(\mathcal{P}_\theta f)}(k/L)}{2\pi\imath k/L}e^{2\pi\mathrm{i}tk/L}\right| &\le \frac{C}{n^q}\sum_{0<|k|\le n/2-1}\frac{1}{|k|}\\
&\le C\frac{\log(n)}{n^q}, \qquad (220)
\end{aligned}
$$

and the tail can be bounded

$$
\begin{aligned}
\left|\sum_{|k|\ge n/2}\frac{\widehat{(\mathcal{P}_\theta f)}(k/L)}{2\pi\imath k}e^{2\pi\mathrm{i}tk/L}\right| &= \left|\sum_{|k|\ge n/2}\frac{\widehat{f}((k/L)\cos(\theta),(k/L)\sin(\theta))}{2\pi\imath k}e^{2\pi\mathrm{i}tk/L}\right|\\
&\le C\sum_{|k|\ge n/2}\frac{1}{k^{q+1}}\\
&\le C\frac{1}{n^q}. \qquad (221)
\end{aligned}
$$

Similarly, for all θ,

$$\left|\beta_\theta[0] - (\widehat{\mathcal{V}\mathcal{P}_\theta f})(0)\right| \le \sum_{0<|k|\le n/2-1} \left|\beta_\theta[k] - \frac{\widehat{(\mathcal{P}_\theta f)}(k/L)}{2\pi\imath k/L}\right| + \sum_{|k|\ge n/2} \left|\frac{\widehat{(\mathcal{P}_\theta f)}(k/L)}{2\pi\imath k/L}\right| \le C\frac{\log(n)}{n^q}. \tag{222}$$

It then follows that for all t and θ,

$$|\nu_x(t,\theta) - (\mathcal{V}\mathcal{P}_\theta f)(t)| \le C\frac{\log(n)}{n^q} \le \frac{C}{n^r}, \tag{223}$$

where C does not depend on θ, t or n.

□

We now complete the proof of Proposition 4.11. For brevity, let $G(y,\theta) = (\mathcal{V}\mathcal{P}_\theta f)(y)$, i.e.

$$G(y,\theta) = \int_{-R}^{y}\int_{-R}^{R} f(s\cos(\theta)+t\sin(\theta), t\cos(\theta)-s\sin(\theta))\,dt\,ds. \tag{224}$$

Then from Corollary 4.14,

$$\begin{aligned}\left|\mathrm{SV}_p(x) - \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\frac{L}{n}\sum_{j=0}^{n-1}|G(t_j,\theta_\ell)|^p\right)^{1/p}\right| &= \left|\left(\frac{1}{n}\sum_{\ell=0}^{n-1}\frac{L}{n}\sum_{j=0}^{n-1}|\nu_x(t_j,\theta_\ell)|^p\right)^{1/p} - \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\frac{L}{n}\sum_{j=0}^{n-1}|G(t_j,\theta_\ell)|^p\right)^{1/p}\right| \\ &\le \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\frac{L}{n}\sum_{j=0}^{n-1}|\nu_x(t_j,\theta_\ell) - G(t_j,\theta_\ell)|^p\right)^{1/p} \\ &\le \frac{C}{n^r}.\end{aligned} \tag{225}$$

Note that $G(y,\theta)$ is uniformly Lipschitz in y. Indeed,

$$\begin{aligned}|G(y,\theta) - G(z,\theta)| &\le \left|\int_{z}^{y}\int_{-R}^{R} f(s\cos(\theta)+t\sin(\theta), t\cos(\theta)-s\sin(\theta))\,dt\,ds\right| \\ &\le \|f\|_\infty L|z-y|.\end{aligned} \tag{226}$$

Consequently,

$$\begin{aligned}&\left|\left(\frac{1}{n}\sum_{\ell=0}^{n-1}\int_{-R}^{R}|G(t,\theta_\ell)|^p\,dt\right)^{1/p} - \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\frac{L}{n}\sum_{j=0}^{n-1}|G(t_j,\theta_\ell)|^p\right)^{1/p}\right| \\ &= \left|\left(\frac{1}{n}\sum_{\ell=0}^{n-1}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}|G(t,\theta_\ell)|^p\,dt\right)^{1/p} - \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}|G(t_j,\theta_\ell)|^p\,dt\right)^{1/p}\right| \\ &\le \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}|G(t,\theta_\ell) - G(t_j,\theta_\ell)|^p\,dt\right)^{1/p} \\ &\le \|f\|_\infty L\left(\frac{1}{n}\sum_{\ell=0}^{n-1}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}|t-t_j|^p\,dt\right)^{1/p} \\ &\le \|f\|_\infty L^{2+1/p}\frac{1}{n}.\end{aligned} \tag{227}$$

Therefore, from (225),

$$\left|\mathrm{SV}_p(x) - \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\int_{-R}^{R}|G(t,\theta_\ell)|^p\,dt\right)^{1/p}\right| \le \frac{C}{n}, \tag{228}$$

where $C > 0$ is independent of p and n.

Next, note that $G(y, \theta)$ is uniformly Lipschitz in θ: indeed, because f has bounded derivative, for all (s,t) in the disc of radius R and for all $0 \le \theta < \pi$,

$$|f(s,t) - f(s\cos(\theta) + t\sin(\theta), t\cos(\theta) - s\sin(\theta))| \le \||\nabla f|\|_\infty L \sin(\theta/2), \tag{229}$$

and so for any y between $-R$ and R,

$$\begin{aligned} |G(y,0) - G(y,\theta)| &\le \int_{-R}^{y}\int_{-R}^{R} |f(s,t) - f(s\cos(\theta) + t\sin(\theta), t\cos(\theta) - s\sin(\theta))|\, dt\, ds \\ &\le \||\nabla f|\|_\infty L^3 \sin(\theta/2). \end{aligned} \tag{230}$$

Let $I_p(\theta) = \|\mathcal{V}\mathcal{P}_\theta f\|_{L^p} = \|G(\cdot,\theta)\|_{L^p}$. Then for each $\ell = 0, \dots, n-1$, with $\theta_\ell = \pi\ell/n$, and all $0 \le \theta < \pi$,

$$\begin{aligned} \left|I_p(\theta) - I_p(\theta_\ell)\right| &= \left|\|G(\cdot,\theta)\|_{L^p} - \|G(\cdot,\theta_\ell)\|_{L^p}\right| \\ &\le \|G(\cdot,\theta) - G(\cdot,\theta_\ell)\|_{L^p} \\ &= \left(\int_{-R}^{R} |G(t,\theta) - G(t,\theta_\ell)|^p dt\right)^{1/p} \\ &\le \||\nabla f|\|_\infty L^{3+1/p} \sin(\pi/(2n)). \\ &\le C\||\nabla f|\|_\infty L^{3+1/p} \frac{1}{n}. \end{aligned} \tag{231}$$

We then have

$$\begin{aligned} &\left| \left(\frac{1}{\pi}\int_0^\pi \int_{-R}^{R} |G(t,\theta)|^p \, dt\, d\theta\right)^{1/p} - \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\int_{-R}^{R} |G(t,\theta_\ell)|^p\, dt\right)^{1/p} \right| \\ &= \left| \left(\frac{1}{\pi}\sum_{\ell=0}^{n-1}\int_{\theta_\ell}^{\theta_{\ell+1}} \int_{-R}^{R} |G(t,\theta)|^p \, dt\, d\theta\right)^{1/p} - \left(\frac{1}{\pi}\sum_{\ell=0}^{n-1}\int_{\theta_\ell}^{\theta_{\ell+1}}\int_{-R}^{R} |G(t,\theta_\ell)|^p\, dt\, d\theta\right)^{1/p} \right| \\ &= \left| \left(\frac{1}{\pi}\sum_{\ell=0}^{n-1}\int_{\theta_\ell}^{\theta_{\ell+1}} I_p(\theta)^p \, d\theta\right)^{1/p} - \left(\frac{1}{\pi}\sum_{\ell=0}^{n-1}\int_{\theta_\ell}^{\theta_{\ell+1}} I_p(\theta_\ell)^p \, d\theta\right)^{1/p} \right| \\ &\le \left(\frac{1}{\pi}\sum_{\ell=0}^{n-1}\int_{\theta_\ell}^{\theta_{\ell+1}} \left|I_p(\theta) - I_p(\theta_\ell)\right|^p d\theta\right)^{1/p} \\ &= C\||\nabla f|\|_\infty L^{3+1/p}\frac{1}{n}, \end{aligned} \tag{232}$$

that is,

$$\left| \|f\|_{SV^p} - \left(\frac{1}{n}\sum_{\ell=0}^{n-1}\int_{-R}^{R} |G(t,\theta_\ell)|^p\, dt\right)^{1/p} \right| \le C\||\nabla f|\|_\infty L^{3+1/p}\frac{1}{n}. \tag{233}$$

Combined with (228), we get

$$|\|f\|_{SV^p} - \mathrm{SV}_p(x)| \le \frac{C}{n}, \tag{234}$$

where C does not depend on p or n. The result for $p = \infty$ then follows by taking the limit as $p \to \infty$.

4.2.2 Proof of Proposition 4.12

First, we note that the third part of the Proposition (almost sure convergence) follows immediately from the second part (concentration bound) by using Lemma 4.7, as in the 1D case.

Recall that the vectors α_θ and β_θ are defined as follows:

$$\alpha_\theta[k] = \frac{L^2}{n^2}\sum_{i,j} Z[i,j]e^{-2\pi \mathrm{i}k(t_i\cos(\theta)+t_j\sin(\theta))/L}, \quad -n/2 \le k \le n/2-1, \tag{235}$$

$$\beta_\theta[k] = \frac{\alpha_\theta[k]}{2\pi \mathrm{i}k/L}, \quad -n/2 < k < n/2, \tag{236}$$

and when $k = 0$,

$$\beta_\theta[0] = -\sum_{0<|k|<n/2} \beta_\theta[k](-1)^k. \tag{237}$$

Define the random vector W by $W[j,\ell] = \nu_Z(t_j,\theta_\ell)$, that is,

$$W[j,\ell] = \frac{1}{L}\sum_{k=-n/2+1}^{n/2-1} \beta_{\theta_\ell}[k]e^{2\pi \mathrm{i}t_j k/L} \tag{238}$$

for $0 \le j \le n-1$. Then for each $1 \le p < \infty$,

$$\mathrm{SV}_p(Z_n) = \left(\frac{L}{n^2}\sum_{\ell=0}^{n-1}\sum_{j=0}^{n-1} |W[j,\ell]|^p\right)^{1/p}, \tag{239}$$

and $\mathrm{SV}_\infty(Z_n) = \|W\|_\infty$.

We define the auxiliary vector X by

$$X[j,\ell] = \sum_{0<|k|<n/2} \beta_{\theta_\ell}[k]e^{2\pi \mathrm{i}t_j k/L}. \tag{240}$$

Note that $W[j,\ell] = (X[j,\ell] + \beta_{\theta_\ell}[0])/L$.

Expectation of $\mathrm{SV}_p(Z_n)$**,** $1 \le p \le 2$**.** For fixed $0 \le i,\ell \le n-1$,

$$\begin{aligned}
X[i,\ell] &= \sum_{0<|k|<n/2} \beta_{\theta_\ell}[k]e^{2\pi \mathrm{i}t_i k/L} \\
&= \frac{L}{2\pi \mathrm{i}}\sum_{0<|k|<n/2} \frac{\alpha_{\theta_\ell}[k]}{k}e^{2\pi \mathrm{i}t_i k/L} \\
&= \frac{L^3}{2\pi \mathrm{i}n^2}\sum_{0<|k|<n/2}\sum_{j,j'} Z[j,j']\frac{1}{k}e^{-2\pi \mathrm{i}k(t_j\cos(\theta_\ell)+t_{j'}\sin(\theta_\ell))/L}e^{2\pi \mathrm{i}t_i k/L} \\
&= \frac{L^3}{2\pi \mathrm{i}n^2}\sum_{j,j'} Z[j,j']\sum_{k=1}^{n/2-1}\frac{1}{k}\sin(k2\pi(t_j\cos(\theta_\ell)+t_{j'}\sin(\theta_\ell)-t_i)/L).
\end{aligned} \tag{241}$$

Therefore, since the $Z[j,j']$ are independent, and using Lemma 4.8,

$$\mathbb{E}\left[X[i,\ell]^2\right] \le CL^6\frac{1}{n^4}\sum_{j=0}^{n-1}\sigma_{jj'}^2 \le CL^6\frac{\sigma^2}{n^2}. \tag{242}$$

A nearly identical argument shows

$$\mathbb{E}\left[|\beta_\theta[0]|^2\right] \le CL^6\frac{\sigma^2}{n^2} \tag{243}$$

as well, for any θ. Since $W[j,\ell] = (X[j,\ell] + \beta_{\theta_\ell}[0])/L$, it then follows that

$$\mathbb{E}\left[W[j,\ell]^2\right] \leq CL^4 \frac{\sigma^2}{n^2}. \tag{244}$$

Consequently,

$$\mathbb{E}[\mathrm{SV}_2(Z_n)^2] = \frac{L}{n^2} \sum_{j=0}^{n-1} \sum_{\ell=0}^{n-1} \mathbb{E}[W[j,\ell]^2] \leq CL^5 \frac{\sigma^2}{n^2}, \tag{245}$$

and so by Jensen's inequality,

$$\mathbb{E}[\mathrm{SV}_2(Z_n)] \leq \sqrt{\mathbb{E}[\mathrm{SV}_2(Z_n)^2]} \leq CL^{5/2} \frac{\sigma}{n}. \tag{246}$$

If $1 \leq p \leq 2$, then $\mathrm{SV}_p(Z_n) \leq L^{1/p-1/2}\mathrm{SV}_2(Z_n)$, and so

$$\mathbb{E}[\mathrm{SV}_p(Z_n)] \leq CL^{2+1/p} \frac{\sigma}{n}. \tag{247}$$

Expectation of $\mathrm{SV}_p(Z_n)$**,** $2 < p \leq \infty$**.** Using the bound (242) on the variance of each $X[j,\ell]$, an identical argument to (171) shows that for each $s > 0$,

$$e^{\mathbb{E}[s\|X\|_\infty]} \leq 2n^2 e^{Cs^2L^6\sigma^2/n^2} \tag{248}$$

and taking $s = \sigma^{-1}L^{-3}n\sqrt{\log(n)}$ gives the bound

$$\mathbb{E}\left[\|X\|_\infty\right] \leq CL^3\sigma \frac{\sqrt{\log(n)}}{n}. \tag{249}$$

It is straightforward to show the same bound for each $\beta_{\theta_\ell}[0]$, and since $W[j,\ell] = (X[j,\ell]+\beta_{\theta_\ell}[0])/L$ and $\mathrm{SV}_\infty(Z_n) = \|W\|_\infty$, it then follows that

$$\mathbb{E}\left[\mathrm{SV}_\infty(Z_n)\right] \leq CL^2\sigma \frac{\sqrt{\log(n)}}{n}. \tag{250}$$

For any $2 < p < \infty$, as in (175),

$$\mathrm{SV}_p(Z_n) \leq \mathrm{SV}_2(Z_n)^{2/p} \cdot \mathrm{SV}_\infty(Z_n)^{1-2/p}, \tag{251}$$

and so, just as in the 1D case, with $r = p/2$ and $s = (1-2/p)^{-1}$, by Hölder's inequality we have

$$\begin{aligned}
\mathbb{E}\left[\mathrm{SV}_p(Z_n)\right] &\leq \mathbb{E}\left[\mathrm{SV}_2(Z_n)^{2/p} \cdot \mathrm{SV}_\infty(Z_n)^{1-2/p}\right] \\
&\leq \mathbb{E}\left[\mathrm{SV}_2(Z_n)^{2r/p}\right]^{1/r} \cdot \mathbb{E}\left[\mathrm{SV}_\infty(Z_n)^{s(1-2/p)}\right]^{1/s} \\
&= \mathbb{E}\left[\mathrm{SV}_2(Z_n)\right]^{2/p} \cdot \mathbb{E}\left[\mathrm{SV}_\infty(Z_n)\right]^{1-2/p} \\
&= CL^{2+1/p} \frac{\sigma \log(n)^{1/2-1/p}}{\sqrt{n}}.
\end{aligned} \tag{252}$$

Concentration of $\mathrm{SV}_p(Z_n)$**,** $1 \leq p \leq 2$**.** The first lemma is a standard result about the concentration of averages of subexponential random variables. A proof is included for the reader's convenience.

Lemma 4.15. *Suppose* $R_1, R_2, \ldots, R_m$ *are (not necessarily independent) subexponential random variables, that is, there are positive* $\tau_1, \ldots, \tau_m$ *such that for all* $t > 0$*,*

$$\mathbb{P}\{|R_k| \geq t\} \leq 2\exp\{-t/\tau_k\}. \tag{253}$$

Let $R = \sum_{k=1}^m R_k/m$*. Then* R *satisfies the tail bound*

$$\mathbb{P}\{|R| \geq t\} \leq 2\exp\{-Ct/\tau\}, \tag{254}$$

where $\tau = \sum_{k=1}^m \tau_k/m$ *and* $C > 0$ *is a universal constant.*

Proof. By Proposition 2.8.1 in [69], there are universal constants $A, B > 0$ such that for each k,

$$A\tau_k \le \|R_k\|_{\psi_1} \le B\tau_k, \tag{255}$$

where $\|R_k\|_{\psi_1} = \inf\{K \ge 0 : \mathbb{E}[e^{|R_k|/K}] \le 2\}$ is the subexponential Orlicz norm of R_k. Because $\|\cdot\|_{\psi_1}$ is a norm, the triangle inequality implies

$$\|R\|_{\psi_1} \le \frac{1}{m}\sum_{k=1}^{m} \|R_k\|_{\psi_1} \le B\tau. \tag{256}$$

Applying Proposition 2.8.1 from [69] to R gives

$$\mathbb{P}\{|R| \ge t\} \le 2\exp\{-At/\|R\|_{\psi_1}\} \le 2\exp\{-(A/B)t/\tau\}, \tag{257}$$

as claimed.

□

First, observe that for each $k \neq 0$,

$$\begin{aligned} |\alpha_\theta[k]|^2 &= \left(\frac{L^2}{n^2}\sum_{i,i'} Z[i,i']e^{-2\pi \mathrm{i}(t_i k\cos(\theta)+t_{i'}\sin(\theta))/L}\right)\overline{\left(\frac{L^2}{n^2}\sum_{j,j'} Z[j,j']e^{-2\pi \mathrm{i}k(t_j\cos(\theta)+t_{j'}\sin(\theta))/L}\right)} \\ &= \frac{L^4}{n^4}\sum_{i,i'} Z[i,i']^2 + \text{cross-terms} \end{aligned} \tag{258}$$

and taking expectations and using $\mathbb{E}[\text{cross-terms}] = 0$ then gives

$$\mathbb{E}\left[|\alpha_\theta[k]|^2\right] = \frac{L^4}{n^4}\sum_{i,i'}\sigma_{ii'}^2 \le L^4\frac{\sigma^2}{n^2}. \tag{259}$$

Therefore, for $k \neq 0$,

$$\mathbb{E}\left[|\beta_\theta[k]|^2\right] \le CL^6\frac{\sigma^2}{n^2k^2}. \tag{260}$$

For $0 < k < n/2$, let $\varphi_\theta[k] = \mathrm{Re}(\beta_\theta[k])$ and $\psi_\theta[k] = \mathrm{Im}(\beta_\theta[k])$, and let $c_k[j] = \cos(2\pi k t_j/L)$ and $s_k[j] = \sin(2\pi k t_j/L)$. Then all $n-2$ vectors c_k and s_k are pairwise orthogonal, and $\|c_k\|_2 = \|s_k\|_2 = \sqrt{n/2}$. Then

$$X[j,\ell] = 2\sum_{0<k<n/2}\Big[\varphi_{\theta_\ell}[k]c_k[j] - \psi_{\theta_\ell}[k]s_k[j]\Big], \tag{261}$$

and

$$W[j,\ell] = \frac{2}{L}\sum_{0<k<n/2}\Big[\varphi_{\theta_\ell}[k]c_k[j] - \psi_{\theta_\ell}[k]s_k[j]\Big] + \frac{1}{L}\beta_{\theta_\ell}[0]. \tag{262}$$

Note that $\|X[\cdot,\ell]\|_2^2 = 2n(\|\varphi_{\theta_\ell}\|_2^2 + \|\psi_{\theta_\ell}\|_2^2)$.

Define the random vector $V \in \mathbb{R}^n$ by

$$V[\ell]^2 = (2/L)\|\varphi_{\theta_\ell}\|_2^2 + (2/L)\|\psi_{\theta_\ell}\|_2^2 + (1/nL)\beta_{\theta_\ell}[0]^2. \tag{263}$$

Then

$$\mathrm{SV}_2(Z_n)^2 = \frac{1}{n}\sum_{\ell=0}^{n-1} V[\ell]^2. \tag{264}$$

Fix ℓ between 0 and $n-1$. The $n-2$ variables $\varphi_{\theta_\ell}[k]$ and $\psi_{\theta_\ell}[k]$ are pairwise independent Gaussians with zero mean and, by (260), respective variances

$$\tau_{\varphi_{\theta_\ell}}[k] \equiv \mathbb{E}\left[\varphi_{\theta_\ell}[k]^2\right] \leq CL^6 \frac{\sigma^2}{n^2 k^2} \tag{265}$$

and

$$\tau_{\psi_{\theta_\ell}}[k] \equiv \mathbb{E}\left[\psi_{\theta_\ell}[k]^2\right] \leq CL^6 \frac{\sigma^2}{n^2 k^2}. \tag{266}$$

Let $\tau_\ell = (\tau_{\varphi_{\theta_\ell}}, \tau_{\psi_{\theta_\ell}})$. Then

$$\|\tau_\ell\|_\infty \leq CL^6 \frac{\sigma^2}{n^2} \tag{267}$$

and

$$\|\tau_\ell\|_2^2 = \|\tau_{\varphi_{\theta_\ell}}\|_2^2 + \|\tau_{\psi_{\theta_\ell}}\|_2^2 \leq CL^{12} \frac{\sigma^4}{n^4} \sum_{k=1}^{n/2-1} \frac{1}{k^4} \leq CL^{12} \frac{\sigma^4}{n^4}. \tag{268}$$

Furthermore,

$$\mathbb{E}\left[\|\varphi_{\theta_\ell}\|_2^2 + \|\psi_{\theta_\ell}\|_2^2\right] = \sum_{k=0}^{n/2-1} \tau_{\varphi_{\theta_\ell}}[k] + \sum_{k=0}^{n/2-1} \tau_{\psi_{\theta_\ell}}[k] \leq \sum_{k=0}^{n/2-1} CL^6 \frac{\sigma^2}{n^2 k^2} \leq CL^6 \frac{\sigma^2}{n^2}. \tag{269}$$

Let $\Delta_\ell = \|\varphi_{\theta_\ell}\|_2^2 + \|\psi_{\theta_\ell}\|_2^2$. As in the proof of the 1D case, from Lemma 1 in [33], for any $s > 0$,

$$\mathbb{P}\left\{\Delta_\ell - \mathbb{E}[\Delta_\ell] \geq s\right\} \leq \exp\left(-C \min\left\{\frac{s^2}{\|\tau_\ell\|_2^2}, \frac{s}{\|\tau_\ell\|_\infty}\right\}\right) \leq \exp\left(-C \min\left\{\frac{n^4 s^2}{L^{12}\sigma^4}, \frac{n^2 s}{L^6 \sigma^2}\right\}\right), \tag{270}$$

where C is a universal constant. Since $\mathbb{E}[\Delta_\ell] \leq CL^6\sigma^2/n^2$, it then follows that for any $s > 0$,

$$\mathbb{P}\left\{(2/L)\|\varphi_{\theta_\ell}\|_2^2 + (2/L)\|\psi_{\theta_\ell}\|_2^2 \geq s\right\} \leq \exp\left(-C \min\left\{\frac{n^4 s^2}{L^{10}\sigma^4}, \frac{n^2 s}{L^5\sigma^2}\right\}\right), \tag{271}$$

for $n \geq L^{5/2}\sigma/\sqrt{s}$. It is straightforward to prove the same bound for $(1/nL)\beta_{\theta_\ell}[0]^2$; and since $V[\ell]^2 = (2/L)\|\varphi_{\theta_\ell}\|_2^2 + (2/L)\|\psi_{\theta_\ell}\|_2^2 + (1/nL)\beta_{\theta_\ell}[0]^2$,

$$\mathbb{P}\left\{V[\ell]^2 \geq s\right\} \leq 2\exp\left(-C \min\left\{\frac{n^4 s^2}{L^{10}\sigma^4}, \frac{n^2 s}{L^5\sigma^2}\right\}\right) \leq 2\exp\left(-C\frac{n^2 s}{L^5\sigma^2}\right) \tag{272}$$

for $n \geq L^{5/2}\sigma/\sqrt{s}$.

Since $\mathrm{SV}_2(Z_n)^2 = (1/n)\sum_{\ell=0}^{n-1} V[\ell]^2$, Lemma 4.15 gives

$$\mathbb{P}\left\{\mathrm{SV}_2(Z_n)^2 \geq s\right\} \leq 2\exp\left(-C\frac{n^2 s}{L^5\sigma^2}\right) \tag{273}$$

$n \geq L^{5/2}\sigma/\sqrt{s}$, or equivalently, for all $t > 0$,

$$\mathbb{P}\left\{\mathrm{SV}_2(Z_n) \geq t\right\} \leq 2\exp\left(-C\frac{n^2 t^2}{L^5\sigma^2}\right), \tag{274}$$

for all $n \geq L^{5/2}\sigma/t$. When $n < CL^{5/2}\sigma/t$, the right side is bounded below by a positive constant, and so for some sufficiently large $A > 0$, the bound

$$\mathbb{P}\left\{\mathrm{SV}_2(Z_n) \geq t\right\} \leq A\exp\left(-C\frac{n^2 t^2}{L^5\sigma^2}\right) \tag{275}$$

is valid for all $t \geq 0$ and $n \geq 2$.

Since $\mathrm{SV}_p(Z_n) \leq L^{1/p-1/2}\mathrm{SV}_2(Z_n)$ for all $1 \leq p \leq 2$, it also follows that

$$\mathbb{P}\left\{\mathrm{SV}_p(Z_n) \geq t\right\} \leq A\exp\left(-B\frac{n^2 t^2}{L^{4+2/p}\sigma^2}\right), \tag{276}$$

for all $t \geq 0$ and $n \geq 2$.

Concentration of $\mathrm{SV}_p(Z_n)$**,** $2 < p \leq \infty$**.** Using (244), each $W[j,\ell]$ is Gaussian with variance bounded by $CL^4\sigma^2/n^2$, and so standard Gaussian tails bounds (e.g. see Chapter 2 in [72]) imply

$$\mathbb{P}\left(|W[j,\ell]| \geq t\right) \leq 2\exp\left(-C\frac{n^2t^2}{L^4\sigma^2}\right). \tag{277}$$

Since $\mathrm{SV}_\infty(Z_n) = \|W\|_\infty$, by the union bound, therefore,

$$\mathbb{P}\left(\mathrm{SV}_\infty(Z_n) \geq t\right) \leq n^2 \max_{0\leq j,\ell\leq n-1} \mathbb{P}\left(|W[j,\ell]| \geq t\right) \leq 2n^2\exp\left(-C\frac{n^2t^2}{L^4\sigma^2}\right). \tag{278}$$

Furthermore, for all $2 < p < \infty$, since $\mathrm{SV}_p(Z_n) \leq L^{1/p}\mathrm{SV}_\infty(Z_n)$, we have the bound

$$\mathbb{P}\left(\mathrm{SV}_p(Z_n) \geq t\right) \leq 2n^2\exp\left(-C\frac{n^2t^2}{L^{4+2/p}\sigma^2}\right), \tag{279}$$

as desired.

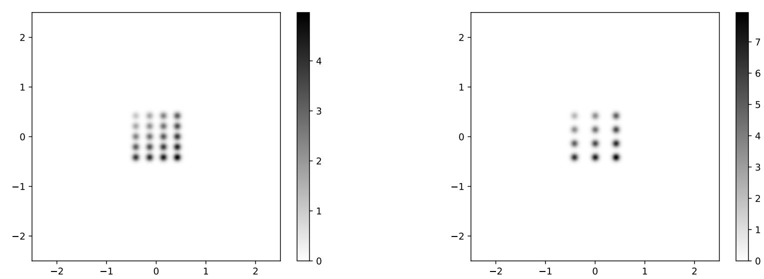

Figure 1: Functions used in the experiment described in Section 5. Left: the source function f. Right: the target function g.

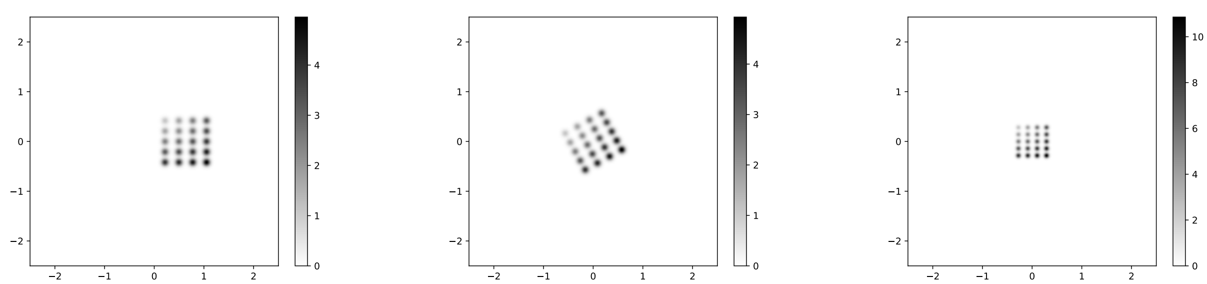

Figure 2: Examples of the deformations applied to the source function f (left panel of Figure 1) in the experiment described in Section 5.1. Left: translation. Middle: rotation. Right: dilation.

5 Numerical results

In this section, we report on numerical experiments that examine the robustness of the sliced Cramér, sliced Wasserstein, and Lebesgue distances for functions of two variables. We focus on two functions, f and g, shown in Figure 1. The function f, shown on the left, will be referred to as the "source" function, and the function g, on the right, will be referred to as the "target" function. Both f and g are convex combinations of isotropic Gaussian functions; f is a combination of 20 Gaussians, while g is a combination of 12 Gaussians. Although f and g are, strictly speaking, not compactly supported, they are numerically supported on a wide enough rectangle $[-r,r]\times[-r,r]$, where in this case we take $r = 2.5$; the width of each Gaussian function comprising f and g is $r/500$.

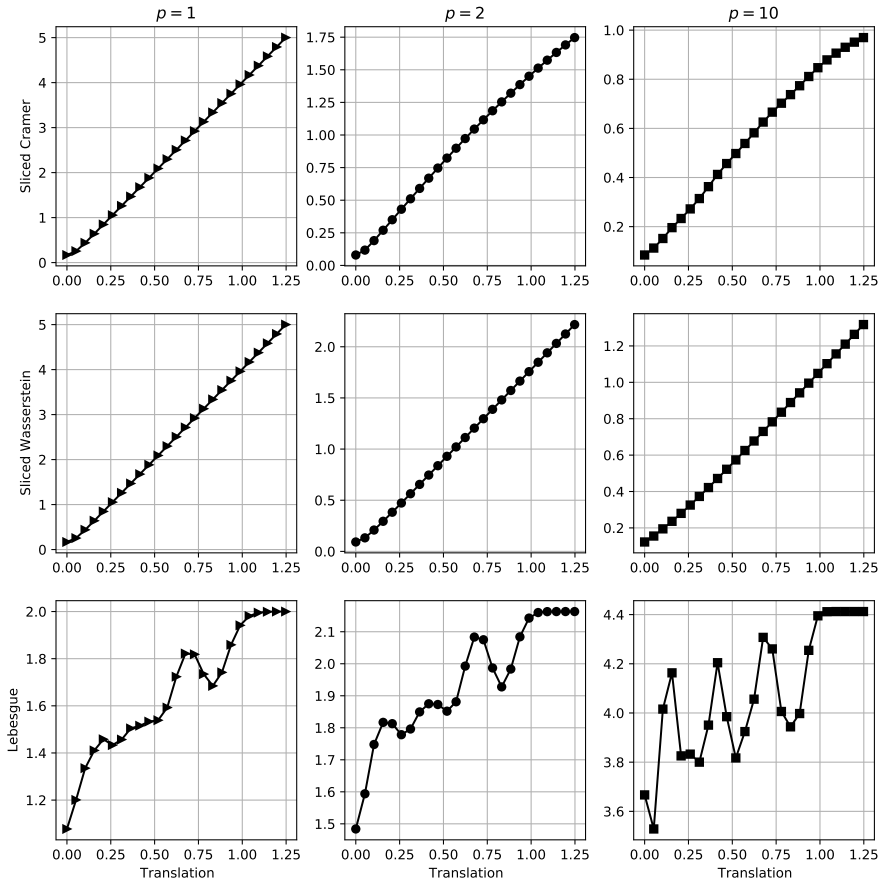

Figure 3: Results of the experiment described in Section 5.1, showing the distances from the translated source function to the target function as a function of the translation size.

The centers of the Gaussians comprising f are arranged in a 5-by-4 grid, whose centers lie equispaced in the rectangle $[-r/6, r/6] \times [-r/6, r/6]$. The centers of the Gaussians comprising g are arranged in a 4-by-3 grid, whose centers also lie equispaced in the rectangle $[-r/6, r/6] \times [-r/6, r/6]$.

We assign non-uniform weights to the Gaussians comprising each function to break rotational symmetry. The weights assigned to the Gaussians in f may be described as follows. Assigning the numbers 1 to 5 to the rows going from top to bottom, and assigning 1 to 4 to the columns going from left to right, the weight assigned to the Gaussian in position (i, j) is proportial to $\sqrt{i^2 + j^2}$. The weights are assigned to the Gaussians in g similarly.

5.1 Robustness to deformations

In our first set of experiments, we examine how the distances between f and g change as f undergoes deformation. For each metric D, we compute the distances $\mathrm{D}(f_{\Phi_\delta}, g)$, where Φ_δ is a deformation depending on a single parameter $\delta \geq 0$, such that $\Phi_0(x) = x$ and the displacement of Φ_δ grows with δ. The distances D are the sliced p-Cramér, sliced p-Wasserstein, and p-Lebesgue, for $p = 1, 2, 10$. We consider three types of deformations: translations, rotations, and dilations. Examples of these are displayed in Figure 2. In all examples, we evaluate the distances for 25 deformation parameters, using samples on a 500-by-500 grid.

Figure 3 shows the distances $\mathrm{D}(f_{\Phi_\delta}, g)$ as a function of the translation size δ, where $\Phi_\delta(x, y) = (x + \delta, y)$. Figure 4 shows the distances $\mathrm{D}(f_{\Phi_\delta}, g)$ as a function of the rotation angle δ, where $\Phi_\delta(x, y) = (x\cos(\delta) - y\sin(\delta), x\sin(\delta) +$

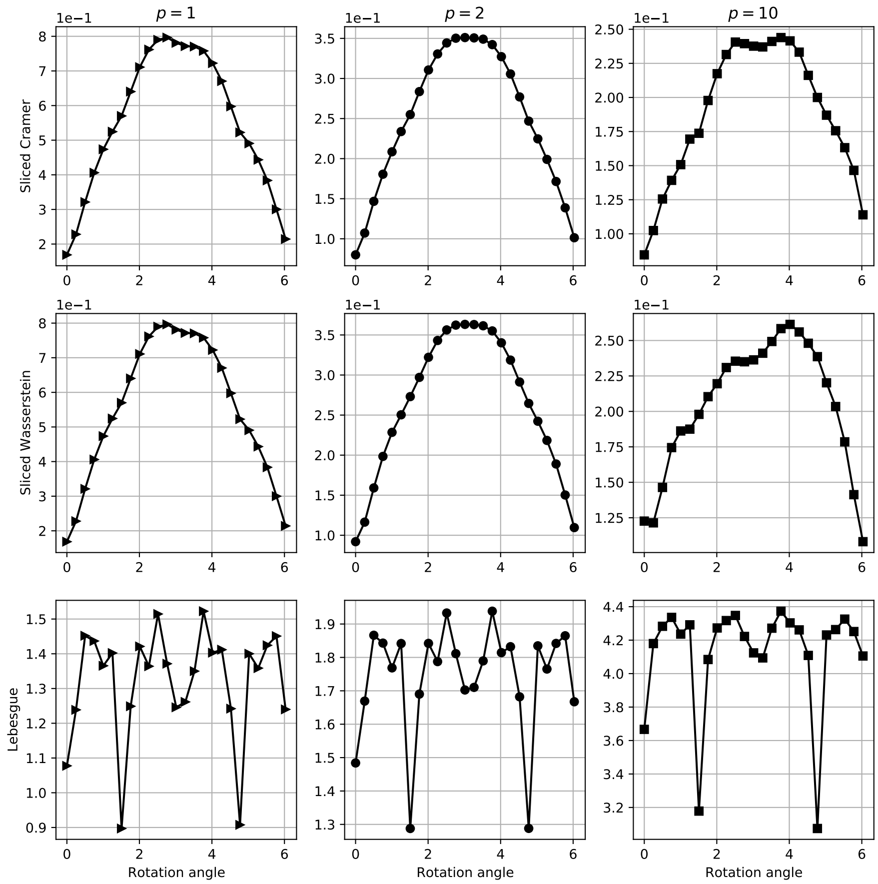

Figure 4: Results of the experiment described in Section 5.1, showing the distances from the rotated source function to the target function as a function of the rotation angle.

$y\cos(\delta))$. Figure 5 shows the distances $\mathrm{D}(f_{\Phi_\delta}, g)$ as a function of the dilation parameter $\delta \geq 1$, where $\Phi_\delta(x) = \delta x$.

From the plots, it is evident that the sliced p-Cramér and sliced p-Wasserstein distances exhibit similar behavior. Both metrics also change more smoothly than the Lebesgue distances as the deformation parameter changes, particularly for translation and rotation. For dilation, the Lebesgue 1-distance quickly becomes large and nearly constant, likely because the supports of the two functions become nearly disjoint as one is dilated. On the other hand, the Lebesgue p-distances for $p > 1$ appear to vary more smoothly with the dilation parameter; this is because the p-norm of the dilated function grows with the dilation size, and so the distance in this case is due to the growing size of the single function, rather than providing any meaningful information about the relationship between the two functions. By contrast, the sliced p-Cramér distance does not grow arbitrarily big as the norm grows; see Remark 7 after the statement of Theorem 3.1.

5.2 Robustness to noise

Next, we examine the robustness of the sliced Cramér distances to additive noise. We consider the functions f and g from before, and add Gaussian noise to the samples of g on a grid of size $n = 512$. We then compute both the sliced Cramér distances and the Euclidean distances between the rotation of f and the noisy samples of g. The noise standard deviations are chosen to be $\sigma = 0.5, 1.0, 1.5$. The distances are averaged over 10 runs of the experiment. The resulting plots are shown in Figure 6. It is evident that the sliced Cramér distances are quite robust to noise at

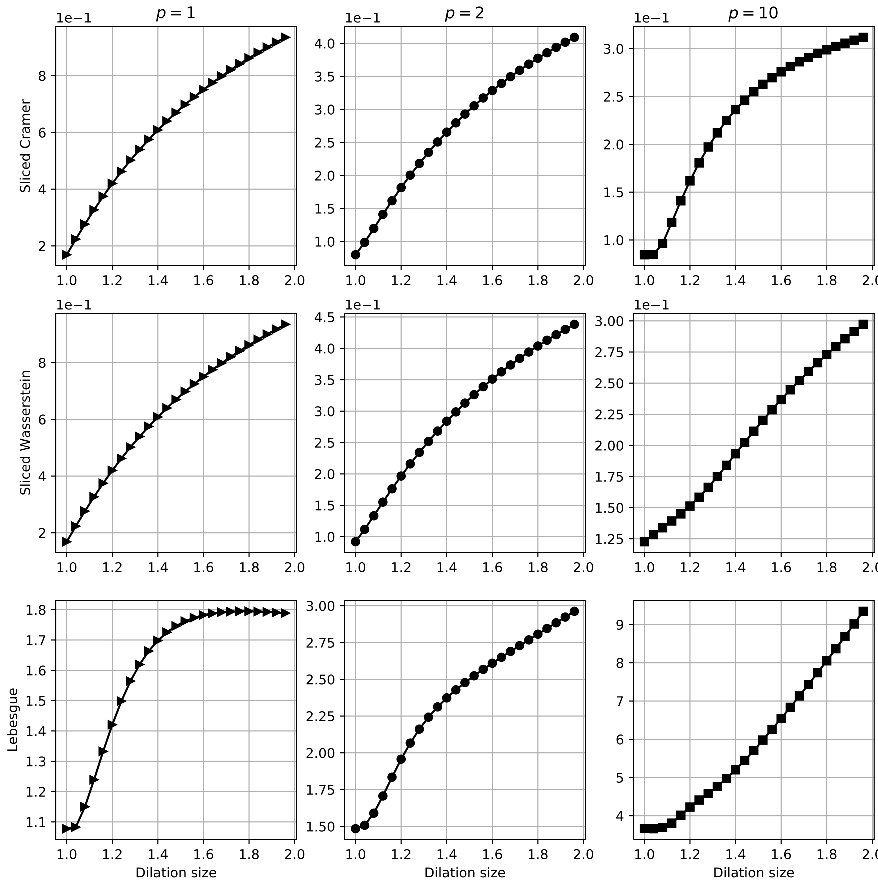

Figure 5: Results of the experiment described in Section 5.1, showing the distances from the dilated source function to the target function as a function of the dilation size.

this sample size, in the sense that though the distances between f and the noisy samples of g are inflated, they still closely track the distances between the noiseless functions (the curve where $\sigma = 0$).

For a more quantitative exploration of the effect of noise on the sliced Cramér distances, we show the following experiment. For increasing values of n, we sample the functions f and g on a grid of size n-by-n, and add Gaussian noise with standard deviation .01 to the samples of g. For $p = 1, 2, 10$, we evaluate the sliced p-Cramér distances between f and the noisy samples of g. For each n, the experiment is repeated $M = 1000$ times. We estimate the true distance d between f and g by evaluating them on a grid of size 2048-by-2048, and measure the average absolute relative error between the noisy distances $\widehat{d_k}$ and d:

$$\mathrm{err}_{n,p} = \frac{1}{M} \sum_{k=1}^{M} \frac{|d - d_k|}{d}. \tag{280}$$

These value are plotted against n^2 in the right panel of Figure 7 (shown in log scale). We also measure the average sliced Volterra p-norms of the noise itself, plotted against n^2 in the left panel of Figure 7. The average errors decay approximately like $O(1/n)$ as n increases (that is, the slopes of the plots are close to $-1/2$), consistent with the error rate established in Theorem 4.10.

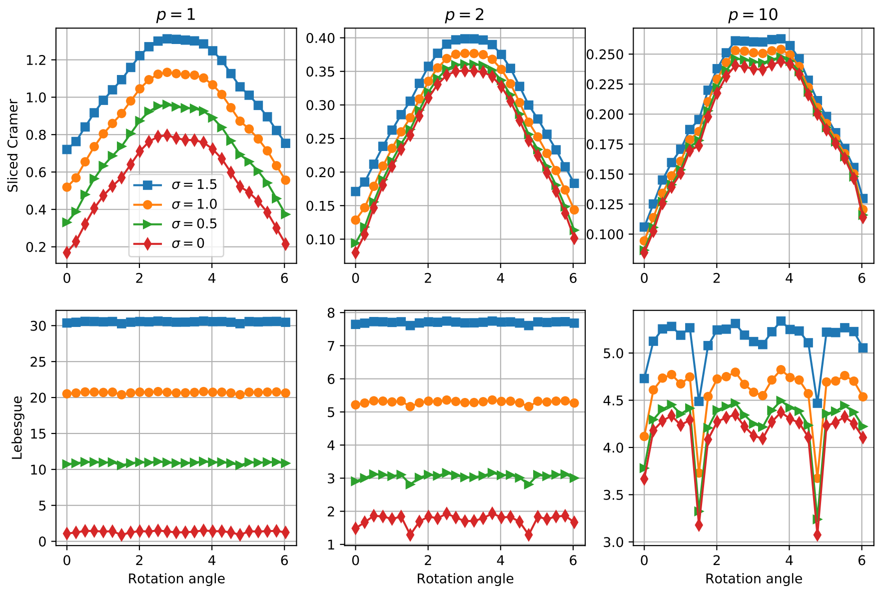

Figure 6: Results of the experiment described in Section 5.2, showing the distances between the rotated source function and the noisy target function as the rotation angle increases and for different noise levels.

6 Conclusion

This paper has proven a number of properties of sliced Cramér distances, showing that these metrics are robust to geometric deformations and noise. Similar geometric robustness properties are shared by Wasserstein distances and their variants; consequently, our results suggest that sliced Cramér distances may be a useful alternative to such distances, especially in applications where the functions being compared are not probability measures or when they are corrupted by noise. In other applications, Wasserstein-type distances may be more appropriate, particularly when the optimal transport (and not just the minimal cost) is of interest.

Because they both exhibit robustness to deformations, it is natural to explore applications of the sliced Cramér distances to problems where Wasserstein and sliced Wasserstein distances have been used previously. For example, one such area is analysis of data from cryo-electron microscopy (cryo-EM), in which one observes two-variable projections of a three-variable volume (a molecule), at unknown viewing directions, from which the volume is to be determined [64, 7, 15]. Wasserstein-type metrics have been proposed for clustering images and parameterizing volumes in cryo-EM [66, 54, 73, 61]. However, the high noise level in cryo-EM images limits the applicability of Wasserstein distances on real data. Because they share many of the same theoretical advantages, it is therefore worthwhile to explore whether sliced Cramér distances are a suitable alternative. Sliced Cramér distances may also be appropriate for heterogeneity analysis in cryo-EM [76, 20, 40, 2, 34, 60, 68, 35], as has been proposed for Wasserstein distances [54].

At the core of the sliced Cramér distances are, of course, the 1D Cramér distances. These are quite simple objects, computed by applying a smoothing filter to the input functions and then evaluating their ordinary Lebesgue distance. It seems likely that one could prove similar robustness results for metrics based on other families of filters. In fact, a number of metrics that have been proposed in recent years are of this type [45, 44, 37, 36, 62, 43]. In future work, we will explore whether such metrics exhibit similar robustness properties as the sliced Cramér and Wasserstein metrics.

Finally, in certain applications one seeks not only robustness to all deformations, but invariance to a specific class of deformations, such as rotations and/or translations [61, 54, 75]. In principle, any metric can be made invariant to a specified set of deformations by simply minimizing the distance over the class of deformations. In some cases of practical interest, such as rigid alignment, this minimization can be done with only small extra computational cost [61, 53]. In the context of the present work, this raises several interesting questions. First, it is known that in the presence of noise, alignment accuracy deteriorates [57, 1]; studying how invariant distances behave under noise, or

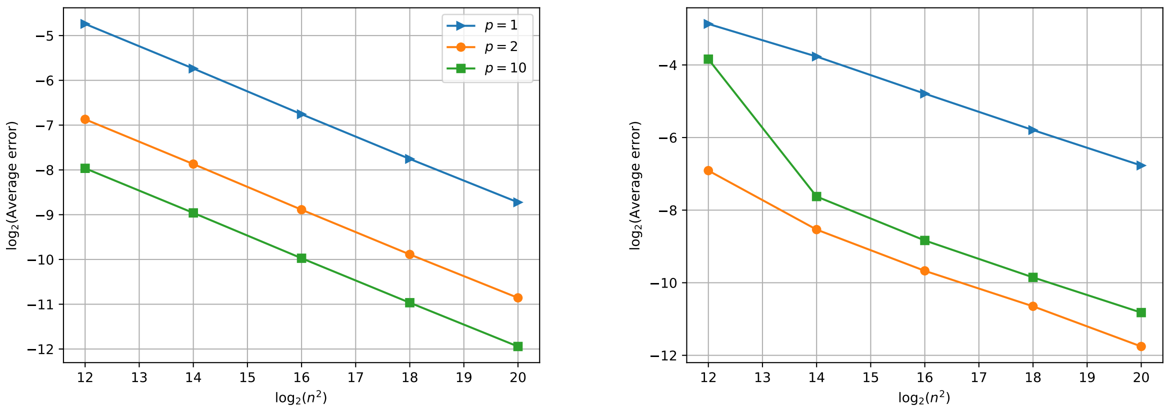

Figure 7: Results of the experiment described in Section 5.2, showing the average errors as a function of number of samples (in log scale). Left: noise-only. Right: signal-plus-noise.

more generally in any setting where alignment cannot be done to high precision, is therefore of interest. Second, a natural question is how introducing invariance to one class of deformations, such as rotations, impacts robustness to other deformations. Questions along these lines will be pursued in future work.

Acknowledgements

I thank Amit Singer for feedback on an earlier version of this work, and Liane Xu for pointing out useful references. I acknowledge support from NSF CAREER award DMS-2238821 and the McKnight Foundation.

References

[1] Cecilia Aguerrebere, Mauricio Delbracio, Alberto Bartesaghi, and Guillermo Sapiro. Fundamental limits in multi-image alignment. *IEEE Transactions on Signal Processing*, 64(21):5707–5722, 2016.

[2] 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.

[3] Joakim Andén and Stéphane Mallat. Deep scattering spectrum. *IEEE Transactions on Signal Processing*, 62(16):4114–4128, 2014.

[4] Alex H. Barnett. Aliasing error of the $\exp(\beta\sqrt{1z^2})$ kernel in the nonuniform fast fourier transform. *Applied and Computational Harmonic Analysis*, 51:1–16, 2021.

[5] Alexander H. Barnett, Jeremy Magland, and Ludvig af Klinteberg. A parallel nonuniform fast fourier transform library based on an "exponential of semicircle" kernel. *SIAM Journal on Scientific Computing*, 41(5):C479–C504, 2019.

[6] Marc G. Bellemare, Ivo Danihelka, Will Dabney, and Shakir Mohamed. The Cramer Distance as a Solution to Biased Wasserstein Gradients. *arXiv preprint arXiv:1705.10743*, 2017.

[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] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Harald Cramér. On the composition of elementary errors. Second paper: Statistical applications. *Scandinavian Actuarial Journal,*, 1:141–180, 1928.

[13] Stanley R. Deans. *The Radon Transform and Some of Its Applications.* 2007.

[14] Ishan Deshpande, Ziyu Zhang, and Alexander G. Schwing. Generative modeling using the sliced Wasserstein distance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3483–3491, 2018.

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

[16] Alok Dutt and Vladimir Rokhlin. Fast Fourier transforms for nonequispaced data. *SIAM Journal on Scientific Computing*, 14(6):1368–1393, 1993.

[17] Alok Dutt and Vladimir Rokhlin. Fast Fourier transforms for nonequispaced data, II. *Applied and Computational Harmonic Analysis*, 2(1):85–100, 1995.

[18] Jeffrey A. Fessler and Bradley P. Sutton. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, 51(2):560–574, 2003.

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

[20] 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.

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

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

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

[24] Leslie Greengard and June-Yub Lee. Accelerating the nonuniform fast Fourier transform. *SIAM Review*, 46(3):443–454, 2004.

[25] Christopher Heil. *Introduction to Real Analysis.* Springer, 2019.

[26] Sigurdur Helgason. *Integral Geometric and Radon Transforms.* 2010.

[27] Gabor T. Herman. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections.* Springer Science & Business Media, 2009.

[28] Long Huang and Dachun Yang. On function spaces with mixed norms – a survey. *Journal of Mathematical Study*, 54(3):262–336, 2021.

[29] Soheil Kolouri, Nicholas A. Ketz, Praveen K. Pilly, and Andrea Soltoggio. Sliced Cramér synaptic consolidation for preserving deeply learned representations. In *ICLR*, 2020.

[30] Soheil Kolouri, Kimia Nadjahi, Shahin Shahrampour, and Umut Şimşekli. Generalized sliced probability metrics. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4513–4517, 2022.

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

[32] 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.

[33] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

[34] 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.

[35] 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.

[36] William Leeb. The mixed Lipschitz space and its dual for tree metrics. *Applied and Computational Harmonic Analysis*, 44(3):584–610, 2018.

[37] 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.

[38] 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.

[39] Alix Lhéritier and Nicolas Bondoux. A Cramér Distance perspective on Quantile Regression based Distributional Reinforcement Learning. In *AISTATS*, 2022.

[40] 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.

[41] 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.

[42] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[43] Gal Mishne, Eric Chi, and Ronald Coifman. Co-manifold learning with missing data. In *International Conference on Machine Learning*, pages 4605–4614, 2019.

[44] Gal Mishne, Ronen Talmon, Israel Cohen, Ronald R. Coifman, and Yuval Kluger. Data-driven tree transforms and metrics. *IEEE Transactions on Signal and Information Processing Over Networks*, 4(3):451–466, 2017.

[45] 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.

[46] 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.

[47] Kimia Nadjahi, Alain Durmus, Lénaïc Chizat, Soheil Kolouri, Shahin Shahrampour, and Umut Şimşekli. Statistical and topological properties of sliced probability divergences.

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

[49] Sloan Nietert, Ziv Goldfeld, Ritwik Sadhu, and Kengo Kato. Statistical, robustness, and computational guarantees for sliced Wasserstein distances. *Advances in Neural Information Processing Systems*, 35:28179–28193, 2022.

[50] Victor M. Panaretos and Yoav Zemel. Statistical aspects of Wasserstein distances. *Annual Review of Statistics and Its Application*, 6:405–431, 2019.

[51] 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.

[52] 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.

[53] Aaditya Rangan, Marina Spivak, Joakim Andén, and Alex Barnett. Factorization of the translation kernel for fast rigid image alignment. *Inverse Problems*, 36(2):024001, 2020.

[54] 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.

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

[56] Maria L. Rizzo and Gábor J. Székely. Energy distance. *Wiley Interdisciplinary Reviews: Computational Statistics*, 8:27–38, 2016.

[57] Dirk Robinson and Peyman Milanfar. Fundamental performance limits in image registration. *IEEE Transactions on Image Processing*, 13(9):1185–1199, 2004.

[58] 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.

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

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

[61] Yunpeng Shi, Amit Singer, and Eric J. Verbeke. Fast alignment of heterogeneous images in sliced Wasserstein distance. *arXiv preprint arXiv:2503.13756*, 2025.

[62] Shirdhonkar Shirdhonkar and David W. Jacobs. Approximate earth mover's distance in linear time. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[63] Albert N. Shiryaev. *Probability–1.* Springer, 3rd edition, 2015. Translated by R.P. Boas and D.M. Chibisov.

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

[65] 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.

[66] Amit Singer and Ruiyi Yang. Alignment of Density Maps in Wasserstein Distance. *Biological Imaging*, 4:E5, 2024.

[67] Gábor J. Székely and Maria L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013.

[68] 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.

[69] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2nd edition, 2025. Free online edition, available at the author's website.

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

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

[72] Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019.

[73] 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.

[74] Ruichong Zhang. Cramer Type Distances for Learning Gaussian Mixture Models by Gradient Descent. *arXiv preprint arXiv:2307.06753*, 2023.

[75] Zhizhen Zhao and Amit Singer. Rotationally invariant image representation for viewing direction classification in cryo-EM. *Journal of structural biology*, 186(1):153–166, 2014.

[76] 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.

[77] Vladimir Mikhailovich Zolotarev. Metric distances in spaces of random variables and their distributions. *Mathematics of the USSR-Sbornik*, 30(3):373–401, 1976.

[78] Vladimir Mikhailovich Zolotarev. Ideal metrics in the problems of probability theory and mathematical statistics. *Australian Journal of Statistics*, 21(3):193–208, 1979.

[79] Vladimir Mikhailovich Zolotarev. Probability metrics. *Theory of Probability and Its Applications*, 28(2):278–302, 1984. Translated by Bernard Seckler.

[80] Antoni Zygmund. *Trigonometric Series*. Cambridge University Press, 3rd edition, 2003.