



Smoothed distance kernels for MMDs and applications in Wasserstein gradient flows

Nicolaj Rux^{1,2} · Michael Quellmalz² · Gabriele Steidl²

Received: 9 April 2025 / Accepted: 29 January 2026
© The Author(s) 2026

Abstract

Negative distance kernels $K(x, y) := -\|x - y\|$ were used in the definition of maximum mean discrepancies (MMDs) in statistics and lead to favorable numerical results in various applications. In particular, so-called slicing techniques for handling high-dimensional kernel summations profit from the simple parameter-free structure of the distance kernel. However, due to its non-smoothness in $x = y$, most of the classical theoretical results, e.g., on Wasserstein gradient flows of the corresponding MMD functional, no longer hold true. In this paper, we propose a new kernel which keeps the favorable properties of the negative distance kernel as being conditionally positive definite of order one with a nearly linear increase towards infinity and a simple slicing structure, but is Lipschitz differentiable now. Our construction is based on a simple 1D smoothing procedure of the absolute value function followed by a Riemann–Liouville fractional integral transform. Numerical results demonstrate that the new kernel performs similarly well as the negative distance kernel in gradient descent methods, but now with theoretical guarantees.

Keywords Negative distance kernel · Maximum mean discrepancy · Conditionally positive definite functions · Wasserstein gradient flows · Optimal transport · Fourier transform

Mathematics Subject Classification (2010) 46E22 · 49Q22 · 42B10 · 44A12 · 65D12

Communicated by: Raymond H. Chan

✉ Nicolaj Rux
nicolaj.rux@math.tu-chemnitz.de
Michael Quellmalz
quellmalz@math.tu-berlin.de
Gabriele Steidl
steidl@math.tu-berlin.de

¹ TU Chemnitz, Faculty of Mathematics, Reichenhainer Straße 39, D-09111 Chemnitz, Germany

² TU Berlin, Institute of Mathematics, Straße des 17. Juni 136, D-10623 Berlin, Germany

1 Introduction

Symmetric, positive definite \mathcal{A} functions have been playing a role in kernel-based learning for a long time [14, 52]. While mostly Gaussian kernels are used, recently, also the conditionally positive definite negative distance kernel $K(x, y) := -\|x - y\|$ has attained interest, e.g., in statistics [50], image dithering/halftoning [16, 21, 33], sampling [39], and generative modeling [25, 30]. Indeed, more general Riesz kernels $K(x, y) := -\|x - y\|^s$, $s \in [0, 2)$, were examined in optimization equilibrium problems (see, e.g., [12, 18, 24]). Let us also mention that gradient flows with respect to the Coulomb kernel $K(x, y) := \|x - y\|^{2-d}$ were quite recently examined in [7] (see also [9]), and $K(x, y) := \|x - y\|^{-1}$ was applied in image halftoning in [48]. For interesting translation invariance properties of MMDs and connections with Wasserstein distances, we refer to [37].

Depending on the kernel, the maximum mean discrepancy (MMD) between two measures can be defined as the sum of an interaction energy and potential energy. Fixing one of the measures, in generative learning called the target measure, Wasserstein gradient flows of the corresponding functional on the Wasserstein-2 space starting in a simple (latent) measure can be applied to sample from that target distribution. While such gradient flows together with numerical forward and backward schemes for their computation are well understood for Lipschitz differentiable kernels (see, e.g., [2, 3]), the convergence behavior of forward steepest descent [27] and Euler backward (JKO) schemes [31] are not clear for the negative distance kernel due to its nondifferentiability in $x = y$. One exception is the one-dimensional case, where the MMD functional becomes, in contrast to higher dimensions, (geodesically) λ -convex (see [15] and the references therein).

Gradient flows of the MMD functional or just the interaction energy with the negative distance kernel or Riesz kernels show a mathematically richer structure than those for smooth kernels and were the object of numerous examinations (see, e.g., [8, 10, 11]). In particular, singular measures can become absolutely continuous along the flow curve and conversely [4, 27], so that these flows are no longer just particle flows when starting in an empirical measure. Finally, let us mention flows in the MMD dissipation geometry [58] which differ from the setting considered in this paper.

If applied in a straightforward way, MMD flows suffer from high computational costs in large-scale computations, since each gradient step requires the computation of kernel sums (or their derivatives) with a large number of summands. For positive definite kernels, a remedy is to apply random Fourier feature techniques [45] based on Bochner's theorem. Unfortunately, the negative distance kernel does not fit into the setting of Bochner's theorem, but here, efficient so-called slicing techniques, which project the high-dimensional problem into a bunch of one-dimensional ones, can be used [26, 28]. For an interesting quite general fast summation approach using deep learning, we refer to [29].

In this paper, we construct a smoothed negative distance kernel such that its MMD functional fulfills the classical assumptions on its Wasserstein gradient flow and ensures in particular that empirical measures evolve as particle flows with proven convergence of Euler forward and backward schemes. On the other hand, these kernels are still conditionally positive definite of order one and behave in applications

similarly to the negative distance kernel, but now with theoretical convergence guarantees.

Our paper is organized as follows: in Section 2, we provide some notation and recall several results on (generalized) Fourier transforms. For readers not familiar with the topic, more material on tempered distributions and the relationship between the generalized and distributional Fourier transforms is added in Appendix 1.

The next three sections contain the steps for defining our smoothed distance kernels: Section 3 starts with appropriate smoothings of the absolute value function in \mathbb{R}^1 . Although not directly relevant for our construction, a relation to the often applied Huber function is addressed in Appendix 2. Then, Section 4 establishes smoothed Euclidean norm functions in \mathbb{R}^d , $d \geq 2$, based on Riemann–Liouville integral transforms, which finally lead to our smoothed kernels in Section 5. Using these kernels, we define reproducing kernel Hilbert spaces and MMDs based on kernel mean embeddings in Section 6. Wasserstein gradient flows of our MMDs are considered in Section 7. We add considerations on the geodesic convexity of the MMDs in Appendix 4. Finally, we demonstrate the very good performance of Wasserstein gradient flows of the MMD with our new kernel by numerical examples in Section 8.

All proofs, which are not indicated to be taken directly from the literature, are given in Appendix 3.

2 Preliminaries

The natural numbers including 0 are denoted by $\mathbb{N} := \{0, 1, 2, \dots\}$. Let $\mathcal{C}_b(\mathbb{R}^d)$ be the space of continuous bounded functions $f: \mathbb{R}^d \rightarrow \mathbb{C}$ with norm

$$\|f\|_\infty := \sup_{x \in \mathbb{R}^d} |f(x)|,$$

$\mathcal{C}_0(\mathbb{R}^d)$ the subspace of functions $f: \mathbb{R}^d \rightarrow \mathbb{C}$ vanishing as $\|x\| \rightarrow \infty$, $\mathcal{C}_c(\mathbb{R}^d)$ the subspace of continuous functions with compact support, $\mathcal{C}^n(\mathbb{R}^d)$, $n \in \mathbb{N}$ the space of n -times continuously differentiable functions, and $\mathcal{C}_c^n(\mathbb{R}^d)$ the space of n -times continuously differentiable functions with compact support. For $1 \leq p \leq \infty$, let $L^p(\mathbb{R}^d)$ be the Banach space of all (equivalence class of) measurable functions $f: \mathbb{R}^d \rightarrow \mathbb{C}$ with finite norm $\|f\|_{L^p}$ and $L^p_{\text{loc}}(\mathbb{R}^d)$ the corresponding locally integrable functions.

Further, we denote by $\mathcal{S}(\mathbb{R}^d)$ the space of complex-valued *Schwartz functions*. The *Fourier transform* $\mathcal{F}: \mathcal{S}(\mathbb{R}^d) \rightarrow \mathcal{S}(\mathbb{R}^d)$ is the bijective mapping defined by

$$\hat{\varphi}(\omega) = \mathcal{F}[\varphi](\omega) := \int_{\mathbb{R}^d} e^{-2\pi i(x,\omega)} \varphi(x) \, dx, \quad \omega \in \mathbb{R}^d. \quad (1)$$

The *Fourier transform* can be extended as a mapping $\mathcal{F}: L^1(\mathbb{R}^d) \rightarrow \mathcal{C}_0(\mathbb{R}^d)$. The *convolution function* $f * g$ of two functions f, g on \mathbb{R}^d is defined, if it exists, by

$$(f * g)(x) := \int_{y \in \mathbb{R}^d} f(x - y)g(y) \, dy = \int_{y \in \mathbb{R}^d} f(y)g(x - y) \, dy, \quad x \in \mathbb{R}^d.$$

In particular, if $f, g \in L^1(\mathbb{R}^d)$, then $f * g$ is defined almost everywhere, and it holds the Fourier convolution theorem

$$\mathcal{F}[f * g] = \hat{f} \hat{g}.$$

For $r \in \mathbb{N}$, we define the space

$$\mathcal{S}_r(\mathbb{R}^d) := \{\varphi \in \mathcal{S}(\mathbb{R}^d) : \varphi(x) \in \mathcal{O}(\|x\|^r) \text{ as } \|x\| \rightarrow 0\}.$$

A measurable function $\hat{f} \in L^2_{\text{loc}}(\mathbb{R}^d \setminus \{0\})$ is called *generalized Fourier transform* of a slowly increasing function $f \in \mathcal{C}(\mathbb{R}^d)$, if there exists an integer $r \in \mathbb{N}$ such that

$$\int_{\mathbb{R}^d} f(x)\hat{\varphi}(x) \, dx = \int_{\mathbb{R}^d} \hat{f}(\omega)\varphi(\omega) \, d\omega \quad \text{for all } \varphi \in \mathcal{S}_{2r}(\mathbb{R}^d), \tag{2}$$

(see [57, Def. 8.9]). If f fulfills (2) for some $r \in \mathbb{N}$, then it also fulfills this relation for all integers larger than r . In particular, if $f \in \mathcal{S}(\mathbb{R}^d)$, then (2) holds for all $r \geq 0$. The smallest $r \in \mathbb{N}$ such that (2) is fulfilled is called *order of the generalized Fourier transform*. We have that \hat{f} is uniquely determined. The generalized Fourier transform differs from the Fourier transform of so-called tempered distributions, in particular of continuous, slowly increasing functions, but coincides with it if restricted to test functions in $\mathcal{S}_{2r}(\mathbb{R}^d)$. This is briefly explained in Appendix 1.

In this paper, we are mainly concerned with powers of the Euclidean norm.

Theorem 2.1 ([57, Thm. 8.16]) *The function $f(x) := \|x\|^\beta, x \in \mathbb{R}^d$, with $\beta > 0, \beta \notin 2\mathbb{N}$ has the generalized Fourier transform*

$$\hat{f}(\omega) = \frac{\Gamma(\frac{d+\beta}{2})}{\pi^{\beta+\frac{d}{2}} \Gamma(-\frac{\beta}{2})} \|\omega\|^{-\beta-d}, \quad \omega \in \mathbb{R}^d$$

of order $r = \lceil \frac{\beta}{2} \rceil$. In particular, we have for $\text{abs}(x) = |x|, x \in \mathbb{R}$, that

$$\widehat{\text{abs}}(\omega) = -\frac{1}{2\pi^2\omega^2}, \quad \omega \in \mathbb{R}. \tag{3}$$

For the generalized Fourier transform, we have the following convolution property.

Proposition 2.2 *Let $f \in \mathcal{C}(\mathbb{R}^d)$ be a slowly increasing function with generalized Fourier transform \hat{f} of order r and $u \in \mathcal{C}_c(\mathbb{R}^d)$. Then, the convolution $f * u \in \mathcal{C}(\mathbb{R}^d)$ is slowly increasing and has a generalized Fourier transform of order r which fulfills $\mathcal{F}[f * u] = \hat{f} \hat{u}$.*

Further, the notation of conditionally positive definiteness will be central in our paper. A continuous, even function $f : \mathbb{R}^d \rightarrow \mathbb{C}$ is *conditionally positive definite of order $r \in \mathbb{N}$* , if for all $N \in \mathbb{N}$, all $x_1, \dots, x_N \in \mathbb{R}^d$, and all $a \in \mathbb{C}^N \setminus \{0\}$ satisfying

$$\sum_{j=1}^N a_j p(x_j) = 0$$

for all d -dimensional polynomials p of degree $\leq r - 1$, we have

$$\sum_{j,k=1}^N a_j \bar{a}_k f(x_j - x_k) \geq 0$$

(see [36, 54]). We denote the space of conditionally positive definite functions of order r by $\text{CP}_r(\mathbb{R}^d)$. In particular, $-\|\cdot\|^\beta \in \text{CP}_1(\mathbb{R}^d)$, $\beta \in (0, 2)$. If $r = 0$, we just speak about positive definite functions. Note that, by this definition, every $f \in \text{CP}_r(\mathbb{R})$ is continuous and even.

Bochner’s theorem characterizes positive definite functions as the Fourier transform of positive measures (see Theorem A.3 in Appendix 1). There are different ways to modify Bochner’s theorem for conditionally positive definite functions. We will use the following one [57, Thm. 8.12].

Theorem 2.3 (Bochner’s theorem for generalized Fourier transform) *Let $f : \mathbb{R}^d \rightarrow \mathbb{C}$ be continuous, slowly increasing, and possess a generalized Fourier transform \hat{f} of order r , which is continuous on $\mathbb{R}^d \setminus \{0\}$. Then, f is conditionally positive definite of order r if and only if \hat{f} is nonnegative.*

Contrary to the generalized Fourier transform of $\|\cdot\|$, its distributional Fourier transform is not a function in the classical sense (see Appendix 1). Together with Bochner’s theorem 2.3, the generalized Fourier transform therefore provides an appropriate framework for studying the (conditional) positive definiteness of the functions in Section 3.

3 Smoothed absolute value function

In this section, we propose to embellish $\text{abs}(x) = |x|$ by convolving it with functions from the set

$$\mathcal{U}^n(\mathbb{R}) := \left\{ u \in C_c^n(\mathbb{R}) : u, \hat{u} \geq 0, u \text{ even}, \int_{\mathbb{R}} u \, dx = 1 \right\}, \quad n \in \mathbb{N}. \quad (4)$$

These functions have the following nice properties.

Proposition 3.1 *Let $u \in \mathcal{U}^n(\mathbb{R})$ and*

$$u_\varepsilon(x) := \frac{1}{\varepsilon} u\left(\frac{x}{\varepsilon}\right), \quad \varepsilon > 0. \quad (5)$$

Then, $f := \text{abs} * u$ fulfills the following:

- i) $f > 0$ and f is even.
- ii) $f(x) = \text{abs}(x)$ for $|x| \geq \text{diam}(\text{supp}(u))/2$.
- iii) $f'' = 2u$ so that f is convex and $f \in C^{n+2}(\mathbb{R})$.
- iv) $-f$ is conditionally positive definite of order $r = 1$, but not positive definite.
- v) $(\text{abs} * u_\varepsilon)(x) = \varepsilon f\left(\frac{x}{\varepsilon}\right)$, $(\text{abs} * u_\varepsilon)'(x) = f'\left(\frac{x}{\varepsilon}\right)$, $(\text{abs} * u_\varepsilon)''(x) = \frac{2}{\varepsilon} u\left(\frac{x}{\varepsilon}\right)$.
- vi) $\text{abs} * u_\varepsilon \rightarrow \text{abs}$ uniformly as $\varepsilon \rightarrow 0$.

The most important functions $u \in \mathcal{U}^n(\mathbb{R})$ in our numerical part will be centered cardinal B -splines. The centered cardinal B -spline of order $m \in \mathbb{N}$, $m \geq 1$, is recursively defined by

$$M_1 := \mathbf{1}_{[-\frac{1}{2}, \frac{1}{2}]}, \quad M_m := M_1 * M_{m-1}, \quad m = 2, 3, \dots$$

B -splines have many useful properties (see [35, 49]).

Proposition 3.2 For the centered cardinal B -splines with $m \geq 1$, the following holds true:

- i) $M_m \geq 0$ and $\int_{\mathbb{R}} M_m(x) \, dx = 1$.
- ii) $\text{supp } M_m = \left[-\frac{m}{2}, \frac{m}{2}\right]$ and M_m is even.
- iii) $M_m \in C^{m-2}(\mathbb{R})$, $m \geq 2$.
- iv) $\widehat{M}_m(\omega) = \text{sinc}^m(\omega)$, where $\text{sinc}(\omega) := \frac{\sin(\pi\omega)}{\pi\omega}$. This is a nonnegative function exactly for even m .
- v) For $m \geq 2$, we have

$$M_m(x) = \frac{1}{(m-1)!} \sum_{k=0}^m (-1)^k \binom{m}{k} \left(x - k + \frac{m}{2}\right)_+^{m-1}, \tag{6}$$

where $x_+ := \max(x, 0)$, and

$$\begin{aligned} M_m(0) &= \frac{2}{\pi} \int_0^\infty \left(\frac{\sin(x)}{x}\right)^m \, dx = \frac{m}{2^{m-1}} \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \frac{(-1)^k (m-2k)^{m-1}}{m! (m-k)!} \\ &= \sqrt{\frac{6}{\pi m}} \left(1 + \mathcal{O}(m^{-1})\right). \end{aligned}$$

- vi) Clearly, it holds $M_{2m} \in \mathcal{U}^{2m-2}(\mathbb{R})$.

The convolution of abs with the centered cardinal B -splines is given in the following proposition.

Corollary 3.3 For $f := \text{abs} * M_m$, it holds

$$f(x) = \frac{2}{(m+1)!} \sum_{k=0}^m (-1)^k \binom{m}{k} \left(x - k + \frac{m}{2}\right)_+^{m+1} - x.$$

Here are two examples.

Example 3.4 From

$$M_2 = \begin{cases} 1 - |x|, & |x| \leq 1, \\ 0, & \text{otherwise,} \end{cases} \quad M_4 = \frac{1}{6} \begin{cases} 3|x|^3 - 6x^2 + 4, & |x| \leq 1, \\ (2 - |x|)^3, & 1 < |x| \leq 2, \\ 0, & \text{otherwise,} \end{cases}$$

we get

$$(\text{abs} * M_2)(x) = \begin{cases} \frac{1}{3}(-|x|^3 + 3|x|^2 + 1), & |x| \leq 1, \\ |x|, & \text{otherwise,} \end{cases} \tag{7}$$

and

$$(\text{abs} * M_4)(x) = \begin{cases} \frac{1}{20}|x|^5 - \frac{1}{6}x^4 + \frac{2}{3}x^2 + \frac{7}{15}, & 0 \leq |x| < 1, \\ \frac{1}{60}(2 - |x|)^5 + |x|, & 1 \leq |x| < 2, \\ |x|, & \text{otherwise.} \end{cases}$$

For a plot of $\text{abs} * M_2$ with its first- and second-order derivatives, see Fig. 1.

Analogously to (5), we write for $m \in \mathbb{N}$ with $m \geq 1$ and $\varepsilon > 0$

$$M_{m,\varepsilon}(x) := \frac{1}{\varepsilon} M_m\left(\frac{x}{\varepsilon}\right), \quad x \in \mathbb{R}.$$

Asking for smoothed absolute value functions, the Huber function may first come into one’s mind. Unfortunately, by the following corollary, the negative Huber function is not conditionally positive definite.

Corollary 3.5 *The Huber function*

$$f(x) := \begin{cases} \frac{1}{2}x^2, & |x| \leq \lambda, \\ \lambda(|x| - \frac{\lambda}{2}), & \text{otherwise,} \end{cases}$$

for $\lambda > 0$ can be rewritten as $f = \lambda (\text{abs} * M_{1,2\lambda}) - \frac{\lambda^2}{2}$ and has the generalized Fourier transform

$$\hat{f}(\omega) = -\frac{\lambda}{2\pi^2\omega^2} \text{sinc}(2\lambda\omega),$$

which takes positive and negative values, so that $-f$ is not conditionally positive definite.

The proof follows from formula (32) in Appendix 2. The Huber function is the so-called Moreau envelope of the absolute value function. Moreau envelopes play an important role in convex analysis. Appendix 2 contains more results on the relation of $\text{abs} * M_m$ to Moreau envelopes, which are interesting on their own.

4 Smoothed Euclidean norm

Our aim is to approximate the Euclidean norm on \mathbb{R}^d by a function which on the one hand keeps its desirable properties, in particular radial symmetry, simple computation, and conditional positive definiteness of order 1, and on the other hand gives rise to Lipschitz differentiable kernels in the next section. First ideas could be the following two:

- Convolve the Euclidean norm in \mathbb{R}^d with some smooth filter. Unfortunately, this is numerically expensive in high dimensions.
- Use $f(\|\cdot\|)$ with $f = \text{abs} * u$ and $u \in \mathcal{U}^n(\mathbb{R})$. Unfortunately, this function is in general not conditionally positive definite, as the following lemma shows.

Lemma 4.1 *For $f = \text{abs} * M_2$, it holds that $-f(\|\cdot\|) \notin \text{CP}_r(\mathbb{R}^d)$ for any $d \geq 2$ and $r \in \mathbb{N}$.*

Since the above approaches do not provide the desired functions, we propose to use the Riemann–Liouville fractional integral transform, which we consider next.

4.1 Riemann–Liouville fractional integral and slicing in \mathbb{R}^d

For $d \in \mathbb{N}, d \geq 2$, the Riemann–Liouville fractional integral $\mathcal{I}_d: L^\infty_{\text{loc}}(\mathbb{R}) \rightarrow C^n(\mathbb{R})$, $n := \lfloor \frac{(d-2)}{2} \rfloor$ is defined by

$$F(s) = \mathcal{I}_d[f](s) := c_d \int_0^1 f(ts)(1-t^2)^{\frac{d-3}{2}} dt \quad \text{for all } s \in \mathbb{R}, \tag{8}$$

where $c_d := \frac{2w_{d-2}}{w_{d-1}}$ and $w_{d-1} := \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}$ denotes the surface area of the sphere \mathbb{S}^{d-1} . For $d = 2$, the term $(1-t^2)^{\frac{d-3}{2}}$ is not bounded, but integrable, so that we require f to be locally bounded in order for (8) to exist. For $d \geq 3$, the term $(1-t^2)^{\frac{d-3}{2}}$ is bounded, and we can define \mathcal{I}_d on $L^1_{\text{loc}}(\mathbb{R})$.

Our approach is motivated by the slicing techniques for fast kernel summation in [26, 28]. In particular, the Riemann–Liouville fractional integral has the following useful property, which relates a high-dimensional radial function to a function on one-dimensional projections of its inputs (see [46]).

Theorem 4.2 *Let $d \in \mathbb{N}, d \geq 2$ and $f \in L^\infty_{\text{loc}}(\mathbb{R})$ be even. Then, the even function $F: \mathbb{R} \rightarrow \mathbb{R}$ defined by the Riemann–Liouville fractional integral (8) fulfills the projection/slicing condition*

$$F(\|x\|) = \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} f(\langle \xi, x \rangle) dx = \mathbb{E}_{\xi \sim \mathcal{U}_{\mathbb{S}^{d-1}}} [f(\langle x, \xi \rangle)], \tag{9}$$

where $\mathcal{U}_{\mathbb{S}^{d-1}}$ denotes the uniform distribution on the sphere. Further, if f is positive definite, then $F(\|\cdot\|)$ is also positive definite for all $d \geq 2$. Conversely, if $F(\|\cdot\|)$ is

positive definite for some $d \geq 2$, then there exists an even positive definite function f on \mathbb{R} such that (8) is fulfilled.

The slicing formula (9) is a special case of the adjoint Radon transform (see [46]). The following two propositions extend the last property of Theorem 4.2 to conditionally positive functions.

Proposition 4.3 *Let $d \in \mathbb{N}$, $d \geq 2$ and $f \in L^\infty_{\text{loc}}(\mathbb{R})$ be even. Further, let $f \in \text{CP}_r(\mathbb{R})$ for $r \in \mathbb{N}$ and $F = \mathcal{I}_d[f]$. Then, $F(\|\cdot\|) \in \text{CP}_r(\mathbb{R}^d)$.*

Proposition 4.4 *Let $d \geq 3$ and let the $\lfloor \frac{d}{2} \rfloor$ -th derivative of $F \in \mathcal{C}^{\lfloor \frac{d}{2} \rfloor}([0, \infty))$ be slowly increasing. Moreover, assume that $F(\|\cdot\|) \in \text{CP}_r(\mathbb{R}^d)$ has a generalized Fourier transform $\rho(\|\cdot\|) \in \mathcal{C}(\mathbb{R}^d \setminus \{0\})$. Then, the function $f \in \text{CP}_r(\mathbb{R})$ with generalized Fourier transform*

$$\hat{f} \in \mathcal{C}(\mathbb{R} \setminus \{0\}), \quad \hat{f}(\omega) = \frac{\omega_{d-1}}{2} \rho(\omega) |\omega|^{d-1},$$

fulfills (8).

4.2 Riemann–Liouville fractional integral of smoothed absolute value

Next, we are interested in the Riemann–Liouville fractional integral of the smoothed absolute value function $f := \text{abs} * u$, $u \in \mathcal{U}^n(\mathbb{R})$. First of all, the absolute value function is an eigenfunction of \mathcal{I}_d (see, e.g., [28]).

Lemma 4.5 *The functions abs^β , $\beta > -1$ are eigenfunctions of \mathcal{I}_d with eigenvalues $\frac{\Gamma(\frac{d}{2})\Gamma(\frac{\beta+1}{2})}{\sqrt{\pi}\Gamma(\frac{d+\beta}{2})}$.*

The Riemann–Liouville fractional integral of $\text{abs} * u$ has the following properties.

Proposition 4.6 *Let $n, d \in \mathbb{N}$ with $d \geq 2$ and $u \in \mathcal{U}^n(\mathbb{R})$. Then, the function $F := \mathcal{I}_d[\text{abs} * u]$ is even, convex, positive, and $(n + 2)$ -times continuously differentiable. Further, it satisfies for $s \rightarrow \infty$ the relation*

$$F(s) = C_d |s| + \mathcal{O}\left(\frac{1}{s}\right), \quad C_d := \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d+1}{2})}.$$

In particular, $F - C_d \text{abs} \in \mathcal{C}_0(\mathbb{R}) \cap L^2(\mathbb{R})$ and $F' \in \mathcal{C}_b(\mathbb{R})$ with $F'(0) = 0$.

The function $F^\varepsilon := \mathcal{I}_d[\text{abs} * u_\varepsilon]$ converges in $L^2(\mathbb{R})$ and also pointwise to $C_d \text{abs}$ as $\varepsilon \rightarrow 0$.

For the special case of B -splines $u := M_m$, we have the following result.

Proposition 4.7 For $m \in \mathbb{N}$ with $m \geq 2$, let $f := \text{abs} * M_m$. Then, we have for $d \geq 2$

$$\mathcal{I}_d[f](s) = c_d \sum_{k=0}^m (-1)^k \binom{m}{k} \sum_{n=0}^{m+1} \frac{(\frac{m}{2} - 2)^{m+1-n}}{n!(m+1-n)!} s^n q_d(n, k - \frac{m}{2}; s) - \frac{\pi c_{d+1}}{2} s, \quad s > 0,$$

where

$$q_d(n, a; s) := \begin{cases} \frac{\Gamma(\frac{d-1}{2})\Gamma(\frac{n+1}{2})}{\Gamma(\frac{d+n}{2})}, & a \leq 0, \\ \frac{\Gamma(\frac{d-1}{2})\Gamma(\frac{n+1}{2})}{\Gamma(\frac{d+n}{2})} - B_{a^2/s^2}(\frac{n+1}{2}, \frac{d-1}{2}), & 0 < a < s, \\ 0, & a \geq s \end{cases}$$

with the incomplete Beta function $B_x(a, b) := \int_0^x t^{a-1}(1-t)^{b-1} dt$ for $a, b > -1$ and $x \in [0, 1]$.

Note that for odd d , the incomplete beta function in $q_d(n, a; s)$ is a polynomial of degree $n + d - 4$ in $1/s$, and hence, $\mathcal{I}_d[f](s)$ is a rational function of $|s|$. In particular, we obtain for $u = M_2$ and $u = M_4$ the following functions F .

Example 4.8 For $d = 3$, it holds

$$\mathcal{I}_3[\text{abs} * M_2](s) = \frac{1}{12} \begin{cases} -|s|^3 + 4s^2 + 4, & |s| \leq 1, \\ 6|s| + \frac{1}{|s|}, & \text{otherwise,} \end{cases}$$

and

$$\mathcal{I}_3[\text{abs} * M_4](s) = \frac{1}{360} \begin{cases} 3|s|^5 - 12s^4 + 80s^2 + 168, & |s| \leq 1, \\ -|s|^5 + 12s^4 - 60|s|^3 + 160s^2 - 60|s| + 192 - \frac{4}{|s|}, & 1 \leq |x| \leq 2, \\ 180|s| + \frac{60}{|s|}, & \text{otherwise.} \end{cases}$$

For an illustration of the first function, see Fig. 1. We have $\mathcal{I}_3[\text{abs} * M_2] \in \mathcal{C}^3(\mathbb{R})$ and $\mathcal{I}_3[\text{abs} * M_4] \in \mathcal{C}^5(\mathbb{R})$.

Based on the previous results, we propose to approximate the negative Euclidean norm on \mathbb{R}^d by

$$\Phi = F(\|\cdot\|) := \mathcal{I}_d[f](\|\cdot\|), \quad f := -\text{abs} * u, \quad u \in \mathcal{U}^n(\mathbb{R}), \quad n \in \mathbb{N}. \quad (10)$$

Summarizing Propositions 3.1 and 4.3, this function has the following properties.

Theorem 4.9 The function Φ in (10) has the following properties:

- i) Φ is conditionally positive definite of order one on \mathbb{R}^d .
- ii) $\Phi(x) < 0$ for all $x \in \mathbb{R}^d$.
- iii) $\Phi(x) = -C_d \|x\| + \varphi(\|x\|)$ with $\varphi \in \mathcal{C}_0(\mathbb{R})$ and $\varphi(s) \in \mathcal{O}(\frac{1}{s})$ as $s \rightarrow \infty$.
- iv) Φ is $n + 2$ times continuously differentiable.
- v) $\nabla \Phi$ is Lipschitz- L continuous with $L := 2\sqrt{d}\|u\|_\infty$

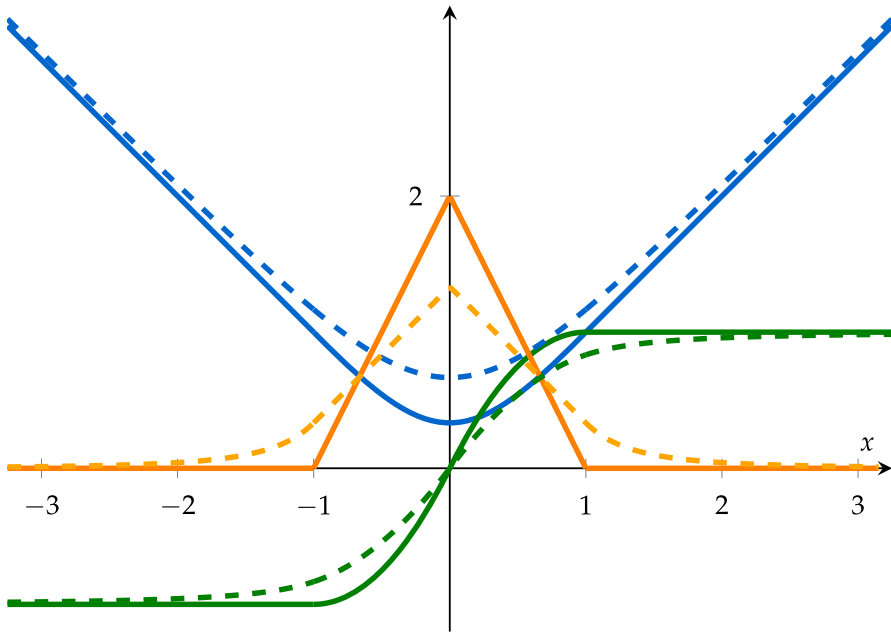


Fig. 1 Smoothed absolute value $f = \text{abs} * M_2$ (solid, blue) with its first (solid, green) and second (solid, orange) derivatives, the latter being equal to $2M_2$, and $F = 2\mathcal{L}_3[f]$ (dashed blue) with its first (dashed, green) and second (dashed, orange) derivatives

vi) Φ is concave and $(-L)$ -convex, i.e., for all $\lambda \in [0, 1]$ and all $x, y \in \mathbb{R}^d$, we have

$$\Phi(\lambda x + (1 - \lambda)y) \leq \lambda\Phi(x) + (1 - \lambda)\Phi(y) + \frac{L}{2}\lambda(1 - \lambda)\|x - y\|^2.$$

5 Smoothed distance kernels

In this section, we show how the above functions Φ induce characteristic kernels with nice Lipschitz properties. These kernels can be used to define MMDs between measures, and the MMDs can then serve as functionals for Wasserstein gradient flows.

We call a symmetric function $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ a *kernel*. A kernel is *positive definite*, if for all $N \in \mathbb{N}$, all $x_1, \dots, x_N \in \mathbb{R}^d$, and all $a \in \mathbb{C}^N$, it holds

$$\sum_{j,k=1}^N a_j a_k K(x_j, x_k) \geq 0.$$

Unfortunately, the kernel $K(x, y) := F(\|x - y\|)$ with F in (10) is not positive definite, since $F(\|\cdot\|)$ is only conditionally positive definite of order $r = 1$. However, we have the following proposition (see [57, Thm 10.18]). Here, $\Pi_{r-1}(\mathbb{R}^d)$ denotes the linear space of d -variate polynomials of degree $\leq r - 1$ which has dimension $N := \binom{d+r-1}{r-1}$.

Proposition 5.1 *Let $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ be a conditionally positive definite function of order $r \in \mathbb{N}$. Let $\Xi := \{\xi_k : k = 1, \dots, N\}$ be a set of points such that $p(\xi_k) = 0$ for all $k = 1, \dots, N$ and any $p \in \Pi_{r-1}(\mathbb{R}^d)$ implies that p is the zero polynomial. Denote by $p_j, j = 1, \dots, N$ the set of Lagrangian basis polynomials with respect to Ξ , i.e., $p_j(\xi_k) = \delta_{j,k}$. Then*

$$\begin{aligned}
 K(x, y) := & \Phi(x - y) - \sum_{j=1}^N p_j(x)\Phi(\xi_j - y) - \sum_{k=1}^N p_k(y)\Phi(x - \xi_k) \\
 & + \sum_{j,k=1}^N p_j(x)p_k(y)\Phi(\xi_j - \xi_k)
 \end{aligned} \tag{11}$$

is a positive definite kernel. In particular, we have in case $r = 1$ that

$$\Phi(x - y) - \Phi(x) - \Phi(y) + \Phi(0)$$

is positive definite, where we can skip the constant third term if $\Phi(0) \leq 0$.

For our kernel from the function in (10), we obtain directly by Theorem 4.9 v) and Proposition 5.1 the following corollary.

Corollary 5.2 *Let $F(\|\cdot\|)$ be defined by (10). Then*

$$K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}, \quad K(x, y) := F(\|x - y\|) - F(\|x\|) - F(\|y\|) \tag{12}$$

is a positive definite kernel and

$$K(x, x) = F(0) - 2F(\|x\|) \in \mathcal{O}(\|x\|). \tag{13}$$

Moreover, K is continuously differentiable with Lipschitz continuous gradient, i.e.,

$$\|\nabla K(x, x') - \nabla K(y, y')\| \leq L(\|x - y\| + \|x' - y'\|) \text{ for all } x, x', y, y' \in \mathbb{R}^d. \tag{14}$$

6 Maximum mean discrepancy with respect to K

A Hilbert space \mathcal{H} of real-valued functions on \mathbb{R}^d is called a *reproducing kernel Hilbert space* (RKHS) if the point evaluations $h \mapsto h(x), h \in \mathcal{H}$, are continuous for all $x \in \mathbb{R}^d$. There exist various textbooks on RKHS from different points of view (see, e.g., [14, 51, 52]). By [52, Thm. 4.20], every RKHS admits a unique positive definite kernel $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, which is determined by the reproducing property

$$h(x) = \langle h, K(x, \cdot) \rangle_{\mathcal{H}} \text{ for all } h \in \mathcal{H}. \tag{15}$$

In particular, we have $K(x, \cdot) \in \mathcal{H}$ for all $x \in \mathbb{R}^d$ and

$$|h(x)| \leq \|h\|_{\mathcal{H}} \|K(x, \cdot)\|_{\mathcal{H}} = \|h\|_{\mathcal{H}} \sqrt{K(x, x)}. \tag{16}$$

Conversely, for any positive definite kernel $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, there exists a unique RKHS with reproducing kernel K , denoted by \mathcal{H}_K [52, Thm. 4.21].

RKHSs are closely related to measure spaces. Let $\mathcal{M}(\mathbb{R}^d)$ denote the space of finite, real-valued Radon measures and $\mathcal{P}(\mathbb{R}^d)$ the space of probability measures on \mathbb{R}^d . Further, let

$$\mathcal{M}_\alpha(\mathbb{R}^d) := \left\{ \mu \in \mathcal{M}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \|x\|^\alpha d\mu(x) < \infty \right\}, \quad 0 < \alpha < \infty$$

and similarly

$$\mathcal{P}_\alpha(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \|x\|^\alpha d\mu(x) < \infty \right\}, \quad 0 < \alpha < \infty.$$

Let $K(x, x) \in \mathcal{O}(\|x\|^\alpha)$. For example, we have by (13) for our kernel in (12) that $\alpha = 1$. Then, it can be seen by (16) that $\mathcal{H}_K \subset L^1(\mu)$ for all $\mu \in \mathcal{M}_{\alpha/2}(\mathbb{R}^d)$ and the so-called *kernel mean embedding* (KME) $m : \mathcal{M}_{\alpha/2}(\mathbb{R}^d) \rightarrow \mathcal{H}_K, \mu \mapsto m_\mu$ given by

$$\langle h, m_\mu \rangle_{\mathcal{H}_K} = \int_{\mathbb{R}^d} h d\mu \quad \text{for all } h \in \mathcal{H}_K \tag{17}$$

is well-defined, meaning that for every $\mu \in \mathcal{M}_{\alpha/2}(\mathbb{R}^d)$, there exists a unique $m_\mu \in \mathcal{H}_K$ such that (17) is fulfilled [52, Lemma 4.24]. In particular, we have by (15) that

$$m_\mu(x) = \int_{\mathbb{R}^d} K(x, y) d\mu(y). \tag{18}$$

The KME is not surjective [53]. For a positive definite kernel K with $K(x, x) \in \mathcal{O}(\|x\|^\alpha)$, the *maximum mean discrepancy* (MMD) $\mathcal{D}_K : \mathcal{M}_{\alpha/2}(\mathbb{R}^d) \times \mathcal{M}_{\alpha/2}(\mathbb{R}^d) \rightarrow \mathbb{R}_{\geq 0}$ is by (16) well-defined by

$$\begin{aligned} \mathcal{D}_K^2(\mu, \nu) &:= \int_{\mathbb{R}^d \times \mathbb{R}^d} K(x, y) d(\mu(x) - \nu(x)) d(\mu(y) - \nu(y)) \tag{19} \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} K(x, y) d\mu(x) d\mu(y) - 2 \int_{\mathbb{R}^d \times \mathbb{R}^d} K(x, y) d\mu(x) d\nu(y) \\ &\quad + \int_{\mathbb{R}^d \times \mathbb{R}^d} K(x, y) d\nu(x) d\nu(y) \\ &= \|m_\mu - m_\nu\|_{\mathcal{H}_K}^2, \end{aligned}$$

(see [6, 23]), where the last equality follows directly from the KME (18). If the KME is injective, then K is called a *characteristic kernel*. In this case, the MMD \mathcal{D}_K is a distance on $\mathcal{M}_{\alpha/2}(\mathbb{R}^d)$. Kernels induced by Gaussians are typical characteristic kernels. By the following proposition, also, our kernel (12) is characteristic, so that \mathcal{D}_K is a distance on $\mathcal{M}_{1/2}(\mathbb{R}^d)$.

Proposition 6.1 *Let K be defined by (12). Then, the kernel mean embedding $m : \mathcal{M}_{1/2}(\mathbb{R}^d) \rightarrow \mathcal{H}_K$ in (17) is injective, i.e., K is a characteristic kernel. More precisely, for all $\mu \in \mathcal{M}_{1/2}(\mathbb{R}^d)$, it holds*

$$\|m_\mu\|_{\mathcal{H}_K}^2 = \frac{1}{\omega_{d-1}\pi^2} \int_{\mathbb{R}^d} |\mu(\mathbb{R}^d) - \hat{\mu}(s)|^2 \frac{\hat{u}(\|s\|)}{\|s\|^{d+1}} ds - F(0) \mu(\mathbb{R}^d)^2, \tag{20}$$

where $\hat{\mu}$ denotes the Fourier transform of μ (see (30)).

In the proof of Proposition 6.1, equation (20) is established first. Then, the localization principle [43, Lem 2.39] implies that the support of \hat{u} is \mathbb{R} , because $u \in \mathcal{U}^n(\mathbb{R})$ is compactly supported. As a consequence, the kernel K is characteristic.

Fortunately, by the following theorem, when dealing with MMDs, it is not necessary to work with the clumsy kernels (11), but instead, we can directly use the conditionally positive definite kernels. Note that the MMD with respect to the negative distance kernel is also known as energy distances in statistics [55].

Theorem 6.2 *Let $\Phi \in \text{CP}_r(\mathbb{R}^d)$ with $r \in \mathbb{N}$, $r \geq 1$ fulfill $\Phi \in \mathcal{O}(\|\cdot\|^\alpha)$, and let $\tilde{K}(x, y) := \Phi(x - y)$. Define the associate positive definite kernel K by (11). Then, $\mathcal{D}_{\tilde{K}}$ in (19) is well-defined for $\mu, \nu \in \mathcal{M}_\alpha(\mathbb{R}^d)$ and \mathcal{D}_K for $\mu, \nu \in \mathcal{M}_\beta(\mathbb{R}^d)$, where $\beta := \max\{r - 1, (r - 1 + \alpha)/2\}$. If $\mu, \nu \in \mathcal{M}_\alpha(\mathbb{R}^d) \cap \mathcal{M}_\beta(\mathbb{R}^d)$ have the same first $r - 1$ moments, i.e.,*

$$\int_{\mathbb{R}^d} p(x) d\mu(x) = \int_{\mathbb{R}^d} p(x) d\nu(x) \quad \text{for all } p \in \Pi_{r-1}(\mathbb{R}^d),$$

then

$$\mathcal{D}_{\tilde{K}}(\mu, \nu) = \mathcal{D}_K(\mu, \nu).$$

For our function $\Phi(x) := F(\|x\|)$ with F in (10), we know already that \mathcal{D}_K is well-defined for measures in $\mathcal{M}_{1/2}(\mathbb{R}^d)$ which is in agreement with the proposition. However, by the proposition, $\mathcal{D}_{\tilde{K}}$ is only well-defined for measures in $\mathcal{M}_1(\mathbb{R}^d)$. If in addition $\int_{\mathbb{R}} d\mu = \int_{\mathbb{R}} d\nu$, then their distances \mathcal{D}_K and $\mathcal{D}_{\tilde{K}}$ are the same. In particular, both distances are well-defined and coincide for measures in $\mathcal{P}_1(\mathbb{R}^d) \supset \mathcal{P}_2(\mathbb{R}^d)$.

By the following remark, there is a relation between the degree of conditional positive definiteness and the growth of a function Φ towards infinity.

Remark 6.3 By [34, Cor 2.3], we have

$$\Phi \in \text{CP}_r(\mathbb{R}^d) \implies \Phi \in \mathcal{O}(\|\cdot\|^{2r}),$$

which implies that $\alpha \leq 2r$ in the assumption of Theorem 6.2. In general, this bound cannot be improved, since $(-1)^r \|\cdot\|^{2r-\varepsilon} \in \text{CP}_r(\mathbb{R}^d)$ for any $r \in \mathbb{N}$, $r \geq 1$ and $\varepsilon \in [0, 2)$ by [57, Cor 8.18] and [54, Lem 3.3]. However, for our function $\Phi(x) := F(\|x\|)$ with F in (10), the above result says that $\Phi \in \mathcal{O}(\|\cdot\|^2)$, but we know already that $\Phi \in \mathcal{O}(\|\cdot\|)$.

Finally, smoothness properties of the kernel transfer to the corresponding RKHS.

Proposition 6.4 For $d \geq 3$ and $n \geq 0$, let $u \in \mathcal{U}^n(\mathbb{R}^d)$. Let the kernel K be given by (12). Then, every $h \in \mathcal{H}_K$ is $\lfloor \frac{n+2}{2} \rfloor$ -times continuously differentiable. If $n \geq 2$ is even, then the gradient ∇h is $\sqrt{2d} \|u''\|_\infty \|h\|_{\mathcal{H}_K}$ Lipschitz continuous.

7 Wasserstein gradient flows of MMDs

7.1 Definition and existence

The behavior of Wasserstein gradient flows of MMDs depends on the kernel in their definition. While there exist many results for smooth kernels like the Gaussian (see, e.g., [3]), gradient flows of MMDs with Riesz kernels and in particular with the negative distance kernel have completely different properties (see, e.g., [27]). In contrast to smooth kernels, empirical measures do in general not remain empirical ones along the flow. Even if a steepest descent scheme, resp. the implicit Euler scheme, exists, a convergence theory is still missing in dimensions larger than one.

Let us briefly recall basic facts on Wasserstein gradient flows (see [2, 47]) and show that our new kernels fulfill all assumptions which are required to ensure the existence of its MMD gradient flow and the convergence of a forward and backward schemes.

For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, we denote by

$$\Pi(\mu, \nu) := \{ \pi \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) : (P_1)_\# \pi = \mu, (P_2)_\# \pi = \nu \}$$

the set of couplings with marginals μ and ν , and by $(P_i)_\# \mu := \mu \circ P_i^{-1} \in \mathcal{P}_2(\mathbb{R}^d)$ the *pushforward* of μ with respect to the projection $P_i(x_1, x_2) := x_i, i = 1, 2$. Together with the Wasserstein distance

$$W_2(\mu, \nu)^2 := \min_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|_2^2 d\pi(x, y), \quad \mu, \nu \in \mathcal{P}_2(\mathbb{R}^d), \quad (21)$$

the set $\mathcal{P}_2(\mathbb{R}^d)$ becomes a complete metric space. The set of optimal couplings in (21) is denoted by $\Pi_{\text{opt}}(\mu, \nu)$. A curve $\gamma : I \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ on an interval $I = [a, b], a < b$ is called *absolutely continuous*, if there exists a Borel velocity field $v : I \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ with $\|v_t\|_{L^2(\mathbb{R}^d; \gamma_t)} \in L^1(I)$ such that the *continuity equation*

$$\partial_t \gamma_t + \nabla_x \cdot (v_t \gamma_t) = 0$$

is fulfilled on $I \times \mathbb{R}^d$ in a weak sense, i.e., for all $\varphi \in C_c^\infty((a, b) \times \mathbb{R}^d)$ it holds

$$\int_0^\infty \int_{\mathbb{R}^d} \partial_t \varphi(t, x) + \langle \nabla_x \varphi(t, x), v_t(x) \rangle d\gamma_t(x) dt = 0.$$

There are many velocity fields corresponding to the same absolutely continuous curve, but only one with minimal $\|v_t\|_{L^2(\mathbb{R}^d; \gamma_t)}$ for a.e. $t \in I$. For a lower semi-continuous

function $G : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, the reduced Fréchet subdifferential ∂G consists of all $v \in L^2(\mathbb{R}^d, \mu; \mathbb{R}^d)$ such that for all $\eta \in \mathcal{P}_2(\mathbb{R}^d)$,

$$G(\eta) - G(\mu) \geq \inf_{\pi \in \Pi_{\text{opt}}(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle v(x), y - x \rangle d\pi(x, y) + o(W_2(\mu, \nu)).$$

If the minimal velocity field in the continuity equation is determined by

$$v_t \in -\partial G(\gamma_t), \quad \text{for a.e. } t > 0, \tag{22}$$

then γ_t is called *Wasserstein gradient flow* of G .

Let $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a characteristic kernel such that its MMD is well-defined for measures in $\mathcal{P}_2(\mathbb{R}^d)$. Examples are Gaussian kernels, the negative distance kernel, and our smoothed negative distance kernels in (12). For a fixed target measure $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, we consider gradient flows of the squared MMD functional

$$G : \mathcal{P}_2(\mathbb{R}^d) \rightarrow [0, \infty), \quad G(\mu) := \frac{1}{2} \mathcal{D}_K^2(\mu, \nu). \tag{23}$$

If K is continuously differentiable, the velocity field in (22) becomes

$$\begin{aligned} v_t &= -\nabla_x \frac{\delta G}{\delta \gamma_t} = -\nabla_x \int_{\mathbb{R}^d} K(\cdot, y) (d\gamma_t(y) - d\nu(y)) \\ &= - \int_{\mathbb{R}^d} \nabla_x K(\cdot, y) (d\gamma_t(y) - d\nu(y)), \end{aligned} \tag{24}$$

(see, e.g., [47]). Here, $\frac{\delta G}{\delta \gamma}$ denotes the functional derivative defined, if it exists, by the function with $\frac{d}{d\epsilon} G(\gamma + \epsilon(\eta - \gamma))|_{\epsilon=0} = \int \frac{\delta G}{\delta \gamma}(\gamma)(d\eta - d\gamma)$ for any $\eta \in \mathcal{P}_2(\mathbb{R}^d)$. Note that $v_t = -\nabla_x m_{\gamma_t - \nu}$ for a positive definite kernel. For the negative distance kernel, we can compute $\frac{\delta G}{\delta \gamma_t}$ as above, but the gradient ∇_x does not exist in $x = y$, i.e., (24) is not well-defined, which causes the different behavior of those flows. The following result guarantees the existence of Wasserstein gradient flows of MMDs with sufficiently smooth kernels and its approximation by an Euler forward scheme.

Proposition 7.1 [3, Prop 1&3] *Let $K \in C^1(\mathbb{R}^d \times \mathbb{R}^d)$ be a positive definite, characteristic kernel that has a Lipschitz continuous gradient in the sense of (14). Then, for any $\nu, \mu \in \mathcal{P}_2(\mathbb{R}^d)$, there exists a unique Wasserstein gradient flow $\gamma : [0, \infty) \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ of the MMD functional (23) starting in $\gamma^{(0)} = \mu$. For a step size $\tau > 0$, we define the Euler forward iteration by*

$$\gamma^{(k+1)} := (I - \tau v^{(k)})_{\#} \gamma^{(k)} \tag{25}$$

where $v^{(k)}$ is related to $\gamma^{(k)}$, $k \in \mathbb{N}$ by (24). The approximated interpolation path

$$\gamma_t^\tau := (I - (t - k\tau)v^{(k)})_{\#} \gamma^{(k)}, \quad t \in [k\tau, (k + 1)\tau),$$

satisfies $W_2(\gamma_t^\tau, \gamma_t) \leq \tau C_T$ for all $t \in [0, T]$, where the constant C_T depends only on $T > 0$.

By Proposition 6.1 and Corollary 5.2, we obtain the following for our smoothed norm kernel.

Corollary 7.2 *The kernel*

$$K(x, y) = F(\|x - y\|) - F(\|x\|) - F(\|y\|)$$

with F from (10) fulfills the conditions of Proposition 7.1. There exists a Wasserstein gradient flow of the corresponding MMD functional (23), and it can be approximated by the Euler forward scheme (25). It holds

$$v_t = - \int_{\mathbb{R}^d} \nabla_x K(\cdot, y) d(\gamma_t - \nu)(y) = - \int_{\mathbb{R}^d} \nabla_x F(\|x - y\|) d(\gamma_t - \nu)(y),$$

so that K can be replaced by $\tilde{K}(x, y) = F(\|x - y\|)$ without changing the flow results.

The last corollary remains valid for $F = \mathcal{I}_{d'}[f]$ with $d' > d$ as follows.

Remark 7.3 Let $d' \geq d$ and $F = \mathcal{I}_{d'}[f]$ for $f \in \text{CP}_r(\mathbb{R})$. By Proposition 4.3, we have $F(\|\cdot\|) \in \text{CP}_r(\mathbb{R}^{d'})$ and hence, also $F(\|\cdot\|) \in \text{CP}_r(\mathbb{R}^d)$. Similarly, if $K_{d'}: \mathbb{R}^{d'} \times \mathbb{R}^{d'} \rightarrow \mathbb{R}$ given by $K_{d'}(x, y) := F(\|x - y\|)$ is a characteristic kernel in $\mathbb{R}^{d'}$, then K_d is also a characteristic in \mathbb{R}^d . Any measure $\mu \in \mathcal{M}_{1/2}(\mathbb{R}^d)$ has the trivial extension $\tilde{\mu} := \mu \otimes \prod_{k=d+1}^{d'} \delta_0 \in \mathcal{M}_{1/2}(\mathbb{R}^{d'})$, where δ_0 is the Dirac measure at 0. Then, the kernel mean embedding (18) satisfies

$$\|m_\mu\|_{\mathcal{H}_{K_d}}^2 = \int_{\mathbb{R}^d \times \mathbb{R}^d} F(\|x - y\|) d\mu(x) d\mu(y) = \int_{\mathbb{R}^{d'} \times \mathbb{R}^{d'}} F(\|x - y\|) d\tilde{\mu}(x) d\tilde{\mu}(y) = \|m_{\tilde{\mu}}\|_{\mathcal{H}_{K_{d'}}}^2.$$

If $\|m_\mu\|_{\mathcal{H}_{K_d}}^2 = 0$, then $\tilde{\mu} = 0$ as $K_{d'}$ is characteristic, and thus $\mu = 0$. Hence, K_d is characteristic. Therefore, we can also use $F = \mathcal{I}_{d'}[f]$ to smooth the negative distance kernel in Corollary 7.2.

There is a more general theory on Wasserstein gradient flows of λ -convex functionals, $\lambda \in \mathbb{R}$, along generalized geodesics (see [2, Thm. 11.2.1]). In Appendix 4, we show that the functional G in (23) with our smoothed negative distance kernel fulfills this λ -convexity with $\lambda < 0$ and establish an analogue to Corollary 7.2 for the Euler backward scheme. In particular, note that it is only ensured for $\lambda > 0$ that the gradient flow converges to the (global) minimizer of G as $t \rightarrow \infty$. Example D.5 in Appendix 4 shows that convergence to the global minimizer ν in (23) is in general not ensured, and the iteration may become stuck in another extreme point.

Finally, let us mention that our new kernel can also be used in the definition of other functionals G , e.g., MMD-regularized f -divergences, where so far only bounded positive definite, characteristic kernels were applied.

Remark 7.4 (MMD-regularized f -Divergence) In [41], inspired by [20], Wasserstein gradient flows of MMD-regularized f -divergences were considered. Unfortunately, the approach requires differentiability of the kernel and therefore does not work for negative distance kernels. In contrast, using Proposition 6.4, it can be shown that our new smoothed distance kernel fits into the setting of the above papers.

7.2 Discretization

In a discrete setting, we consider probability measures

$$\mu := \frac{1}{N} \sum_{n=1}^N \delta_{x_n}, \quad \nu := \frac{1}{M} \sum_{m=1}^M \delta_{y_m}, \quad x_n, y_m \in \mathbb{R}^d,$$

where δ_x is the Dirac measure at $x \in \mathbb{R}^d$. The MMD (19) between these measures is

$$\mathcal{D}_K^2(\mu, \nu) = \frac{1}{N^2} \sum_{n,n'=1}^N K(x_n, x_{n'}) - \frac{2}{MN} \sum_{n,m=1}^{N,M} K(x_n, y_m) + \frac{1}{M^2} \sum_{m,m'=1}^M K(y_m, y_{m'}).$$

The Wasserstein gradient flow of the MMD with a kernel fulfilling the assumptions of Proposition 7.1 keeps the empirical measure structure and moves just the positions of the Dirac measures. Now, let additionally K be radial $K(x, y) = F(\|x - y\|)$ with some even function $F \in \mathcal{C}^2(\mathbb{R})$. Then, the forward Euler scheme (25) reads as

$$\begin{aligned} x_i^{(k+1)} = & x_i^{(k)} - \tau \left(\frac{1}{2N} \sum_{n=1}^N (x_i^{(k)} - x_n^{(k)}) \frac{F'(\|x_i^{(k)} - x_n^{(k)}\|)}{\|x_i^{(k)} - x_n^{(k)}\|} \right. \\ & \left. - \frac{1}{M} \sum_{m=1}^M (x_i^{(k)} - y_m) \frac{F'(\|x_i^{(k)} - y_m\|)}{\|x_i^{(k)} - y_m\|} \right). \end{aligned} \tag{26}$$

Because $F \in \mathcal{C}^2(\mathbb{R})$ is even, we have $F'(0) = 0$ and by L'Hôpital's rule

$$F''(0) = \lim_{s \rightarrow 0} \frac{F'(s)}{s}$$

is well-defined.

For the negative distance kernel $K(x, y) = F(\|x - y\|)$ with $F(s) = -|s|$, Wasserstein gradient flows are not known to exist for dimension $d \geq 2$, because the squared MMD functional G in (23) is not geodesically λ -convex (cf. [27]). However, we can replace $G(\mu)$ by $+\infty$ if μ is not an empirical measure (see, e.g., [30]) and use the Euler scheme (25). Then, the summands in (26) have just the form $\frac{x}{\|x\|}$ with $x \in \{x_i^{(k)} - x_n^{(k)}, x_i^{(k)} - y_m : i, n = 1, \dots, N; m = 1, \dots, M\}$ if $x \neq 0$, and we set $\frac{x}{\|x\|} := 0$ for $x = 0$.

The following Proposition 7.5 is for the flow of a single Dirac, but gives an intuition when $x_i^{(k)}$ is already close to y_i . It shows that with fixed τ , the flow for the negative distance kernel tends to oscillate near the target, while it converges for the smoothed kernel (10). In (26), the repulsion term of x_i and x_j approximately cancels with the attraction term of x_i and y_j , leaving only the attraction of x_i and y_i in form of gradient descent (27) of $-F(\|\cdot - y_i\|)$ on \mathbb{R}^d .

Proposition 7.5 *Let $F \in C^1((0, \infty))$. For the target measure $\nu = \delta_y$ and the initial measure $\gamma^{(0)} = \delta_{x^{(0)}}$ with $y, x^{(0)} \in \mathbb{R}^d$, the sequence $x^{(k)}$ from (26) simplifies to*

$$x^{(k+1)} = x^{(k)} + \tau(x^{(k)} - y) \frac{F'(\|x^{(k)} - y\|)}{\|x^{(k)} - y\|}, \tag{27}$$

and we have the following:

- i) If $F = -\frac{1}{2} \text{abs}$ with step size $\tau > 0$ and $0 < \|y - x^{(0)}\| < \frac{\tau}{2}$, then $x^{(k)} = x^{(0)}$ for even k and $x^{(k)} = x^{(1)}$ for odd k . In particular, $(x^{(k)})_k$ does not converge to y .
- ii) If $F = -\mathcal{I}_d[\text{abs} * u]$ with $u \in \mathcal{U}^0(\mathbb{R})$, then for sufficiently small τ and $\|x^{(0)} - y\| < \tau$, the sequence $(x^{(k)})_k$ converges exponentially to y .

The proof of Proposition 7.5 is given in Appendix 4.

8 Numerical results

We compare the gradient flows (26) of $G = \frac{1}{2} \mathcal{D}_K^2(\cdot, \nu)$ for $K(x, y) = F(\|x - y\|)$ with different functions F . For the first part of the numerics, we use the two-dimensional targets ν as in Fig. 2. In the second part, we use the high-dimensional MNIST dataset. All computations are performed using PyTorch on an Intel i7-10700 CPU with 32 GB of memory and an NVIDIA GeForce RTX 2060 GPU.

8.1 Examples in 2D

For the two-dimensional examples ($d = 2$), we use the following kernels:

- a) **Gaussian** $F(s) := \exp(\frac{-s^2}{2\sigma^2})$ for $\sigma > 0$,
- b) **SND**: smoothed negative distance

$$F := -\mathcal{I}_3[\text{abs} * u_\varepsilon] \text{ with } u_\varepsilon(x) := \frac{1}{\varepsilon} M_2(\frac{x}{\varepsilon}), \tag{28}$$

- c) **ND**: negative distance $F := -\frac{1}{2} \text{abs}$.

The usage of \mathcal{I}_3 instead of \mathcal{I}_2 for the SND kernel is justified by Remark 7.3. The reason is the simple structure of $\mathcal{I}_d[\text{abs} * M_m]$ for $d = 3$ (see Example 4.8), as opposed to $d = 2$. The constant in the ND kernel is chosen so that it is the limit of the SND kernel for $\varepsilon \rightarrow 0$ (see Proposition 4.6).

For the SND kernel, we found choices $\varepsilon \in [10^{-4}, 10^{-2}]$ to work generally well. During testing, we did not encounter numerical issues due to small ε ; instead, the

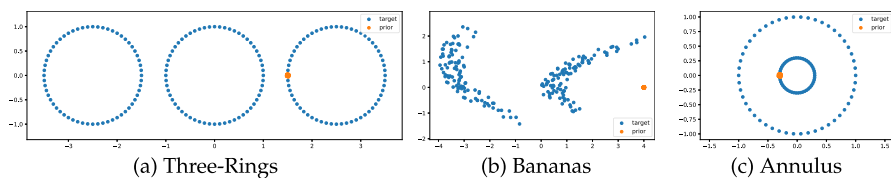


Fig. 2 Target measures ν (blue) and initialization $\gamma^{(0)}$ (orange)

behavior of the SND approaches that of the ND kernel. Large ε over-smooths the kernel, hurting the numerical performance.

8.1.1 Three-rings target

The three-rings target ν in Fig. 2a from [20, Fig. 1] consists of three circles in \mathbb{R}^2 with radius 1 and midpoints $(-2.5, 0)$, $(0, 0)$ and $(2.5, 0)$ discretized with $M = 3 \cdot 40 = 120$ points. The initialization $\gamma^{(0)}$ is a highly localized Gaussian with standard deviation 10^{-4} (see Fig. 2a).

We computed the iteration (25) with step size $\tau = 0.01$ in double precision and display the flow after $k \in \{1\,000, 5\,000, 10\,000, 50\,000\}$ iterations or equivalently after time $t = \tau k$. Figure 3a shows the flows for the Gaussian kernel with standard deviation $\sigma \in \{0.06, 0.3, 1\}$. Here, the quality of the result heavily depends on the choice of σ . If σ is too small, the points cover only two circles; if too large, the points do not lie on the circles. The sweet spot is around $\sigma = 0.3$, but even then, some particles get stuck far from the target. Figure 3b shows the flows for the SND kernel (28) for $\varepsilon \in \{1, 0.1, 0.01\}$. Here, it is preferable to choose a small ε , then all three circles are recovered well. Figure 3c depicts the results with the ND kernel. The flows in Fig. 3c and b for $\varepsilon = 0.01$ are almost identical.

In Fig. 4, we plot the Wasserstein error $W_2(\gamma_t^\tau, \nu)$ between the three-rings target measure ν and the discretized Wasserstein gradient flow γ_t^τ at time t computed with PythonOT [17]. The first plot in Fig. 4 corresponds to the flow γ_t^τ shown in Fig. 3a, b, and c. The remaining plots in Fig. 4 depict the same experiment with different step sizes τ and machine precision, where we always used the same random seed. Regardless of precision, step size, or bandwidth σ , the Gauss kernel stagnates away from the target measure ν .

Proposition 7.5 gives an intuition for the behavior of the ND and SND kernel close to the target. For the ND kernel, Proposition 7.5 i) indicates that μ_t^τ oscillates around the target for any $\tau > 0$ without convergence. In contrast, Proposition 7.5 ii) states that for the SND kernel, $W_2(\mu_t^\tau, \nu)$ decays exponentially if τ is sufficiently small. Numerically, Fig. 4 confirms this behavior. In single precision, ND and SND with $\varepsilon = 0.01$ plateau at $\approx 10^{-3}$. With double precision, ND oscillates at the same error, while SND drops to $\approx 10^{-7}$. Thus, SND matches ND globally but exhibits better local convergence for fixed $\tau > 0$ due to its smoothness.

In Appendix 5, we consider the SND with M_4 instead of M_2 , provide an additional example with two concentric circles, and report computation times.

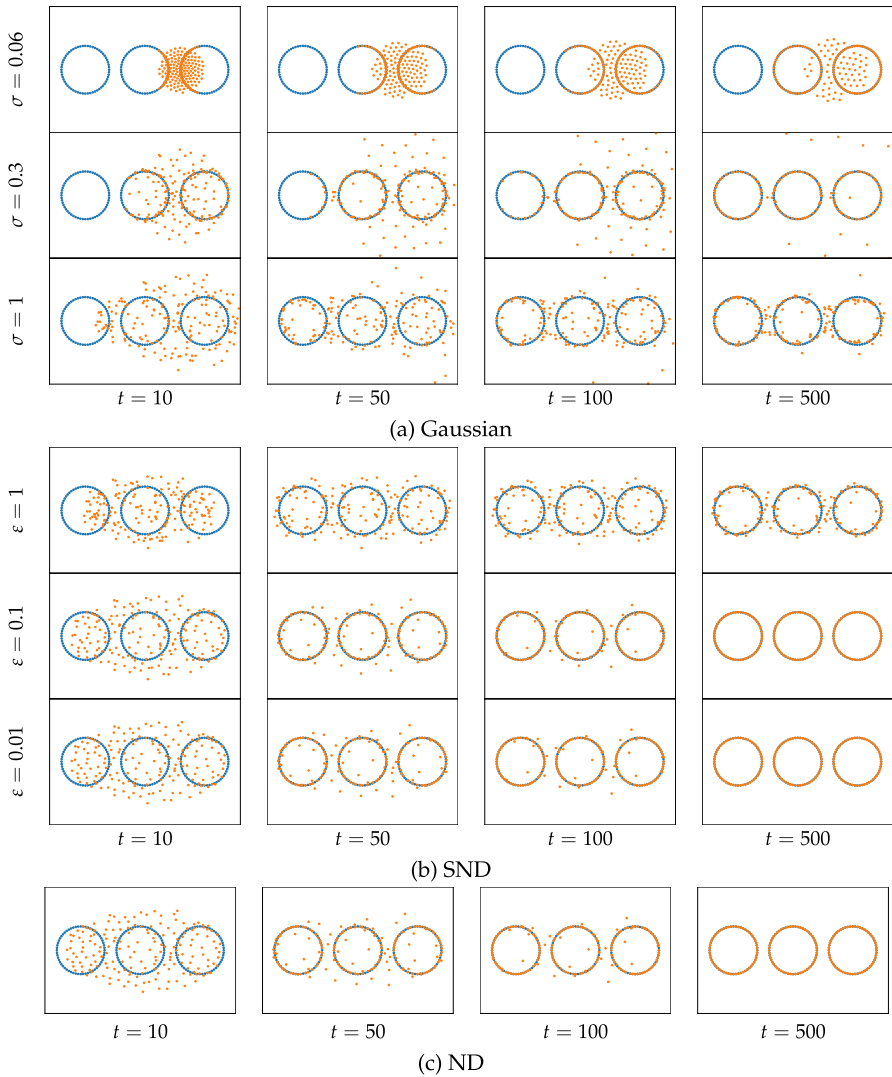


Fig. 3 MMD flow (26) with step size $\tau = 0.01$. For the Gaussian kernel, the result depends heavily on the choice of the parameter σ . For our SND kernel with small ε , the performance is as good as for the ND kernel, which is better than for the Gaussians

8.1.2 Bananas target

The Bananas target ν in Fig. 2b is inspired by the talk from Aude Genevay¹ and its implementation by Viktor Stein². The target consists of two banana shaped clusters in \mathbb{R}^2 , where each banana consists of 100 points, so $M = 200$.

¹ MIFODS Workshop on Learning with Complex Structure 2020, see https://youtu.be/TFdIJib_zEA.

² https://github.com/ViktorAJStein/Regularized_f_Divergence_Particle_Flows.

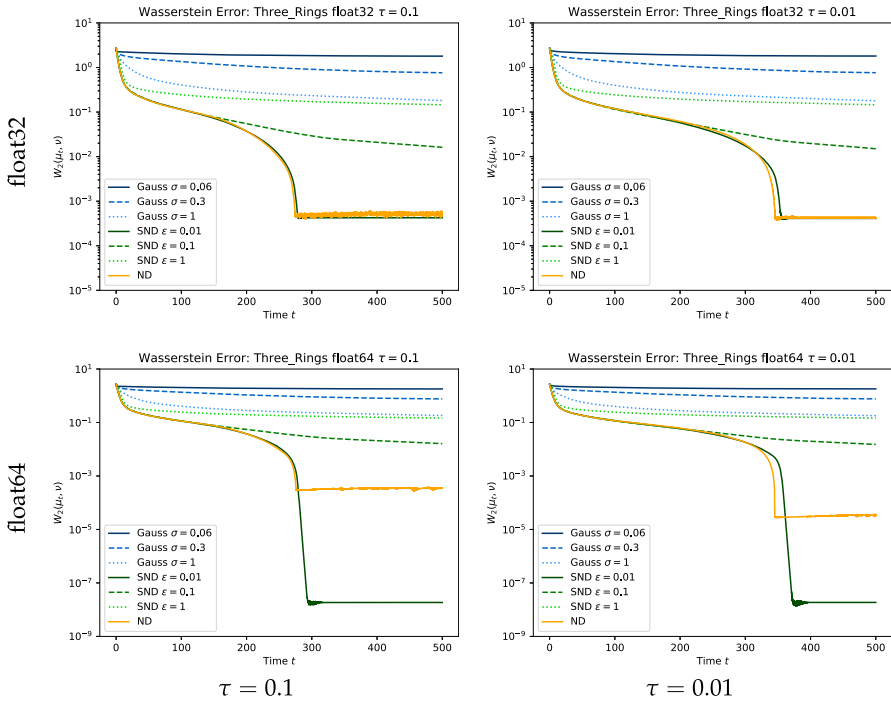


Fig. 4 W_2 error between three-rings target ν and flow γ_t^τ after k with time $t = \tau k$. We compare single precision (first row) and double precision (second row) for step sizes $\tau = 0.1$ (left) and $\tau = 0.01$ (right). In single precision, SND with $\varepsilon = 0.01$ and ND have the smallest error which gets stuck in $\approx 10^{-3}$. In double precision, SND with $\varepsilon = 0.01$ even outperforms ND. For some explanation, see Proposition 7.5

We compute the flows with step size $\tau = 0.02$ in double precision for the Gauss kernel with $\sigma \in \{0.06, 0.3, 1\}$, the SND kernel with $\varepsilon \in \{0.1, 0.01, 0.001\}$, and the ND kernel (see Fig. 5). For small $\sigma = 0.06$, the Gauss kernel struggles to reach the bananas. When $\sigma = 0.3$, the right banana is reached, but some particles blow up and leave the frame. For $\sigma = 1$, the flows reach the bananas, but collapse in the modes and do not recover the structure of the target. In contrast, the SND flow always manages to reach both bananas without blowing up, while a smaller ε again gives more desirable results. The respective Wasserstein errors in Fig. 6 show a similar behavior as for the rings.

8.2 MNIST dataset

We consider the MNIST dataset as our target, where each 28×28 image is considered a point $y \in \mathbb{R}^d$ with $d = 784 = 28^2$. We use $N = M = 100$ images as flow and target.

Fast summation by slicing The computation of (26) includes the summation of kernel values of the form $s_m = \sum_{n=1}^N w_n F'(\|x_n - y_m\|)$ for $m = 1, \dots, M$ with some weights $w_n \in \mathbb{C}$. This summation requires $O(NM)$ arithmetic operations. If $F = \mathcal{I}_d[f]$,

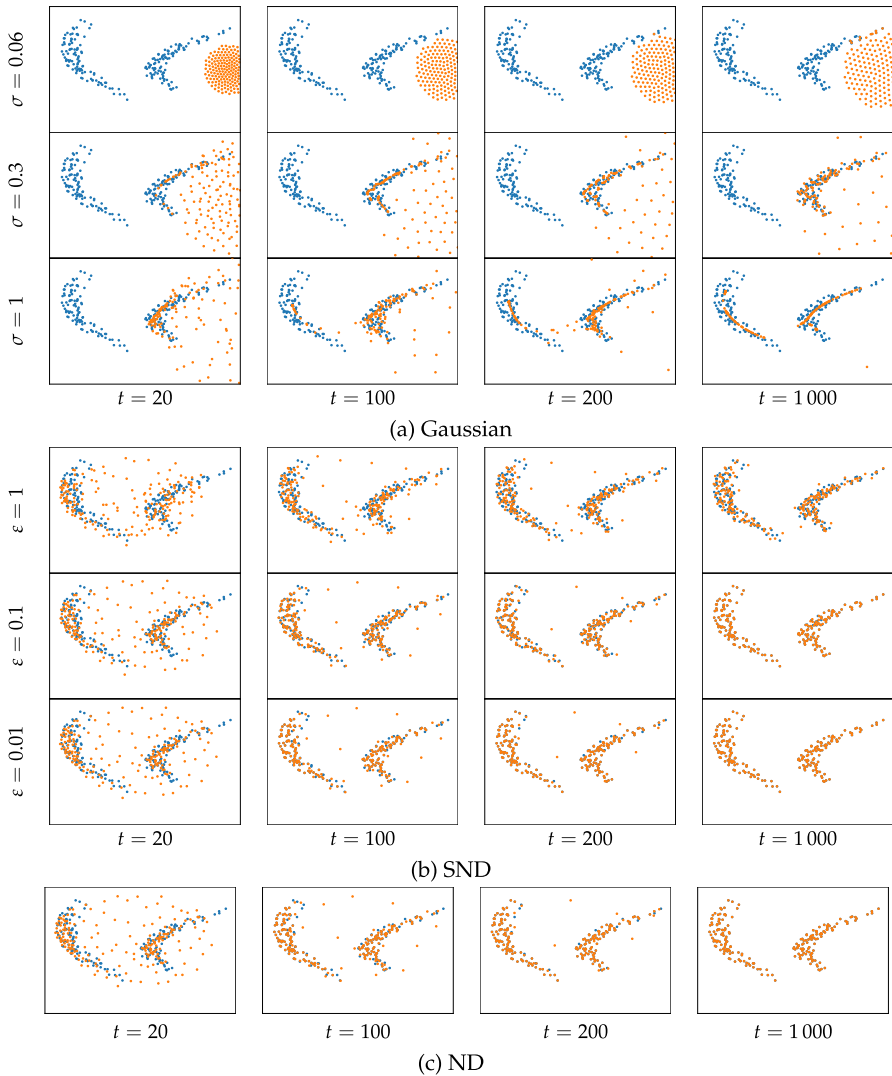


Fig. 5 MMD flow (26) with step size $\tau = 0.02$. For the Gaussian kernel, the result depends heavily on the choice of the parameter σ . For our SND kernel with small ε , the performance is as good as for the ND kernel, which is better than for the Gaussians

then we have by (9) that $F'(x) = \mathbb{E}_{\xi \sim \mathcal{U}_{\mathbb{S}^{d-1}}} [\xi f'(\langle x, \xi \rangle)]$. In order to speed up the computation, the sum s_m can be approximated by slicing [26, 30] via

$$s_m = \mathbb{E}_{\xi \sim \mathcal{U}_{\mathbb{S}^{d-1}}} \left[\sum_{n=1}^N w_n \xi f'(\langle x_n - y_m, \xi \rangle) \right] \approx \frac{1}{P} \sum_{p=1}^P \xi_p \sum_{n=1}^N w_n f'(\langle x_n - y_m, \xi_p \rangle), \tag{29}$$

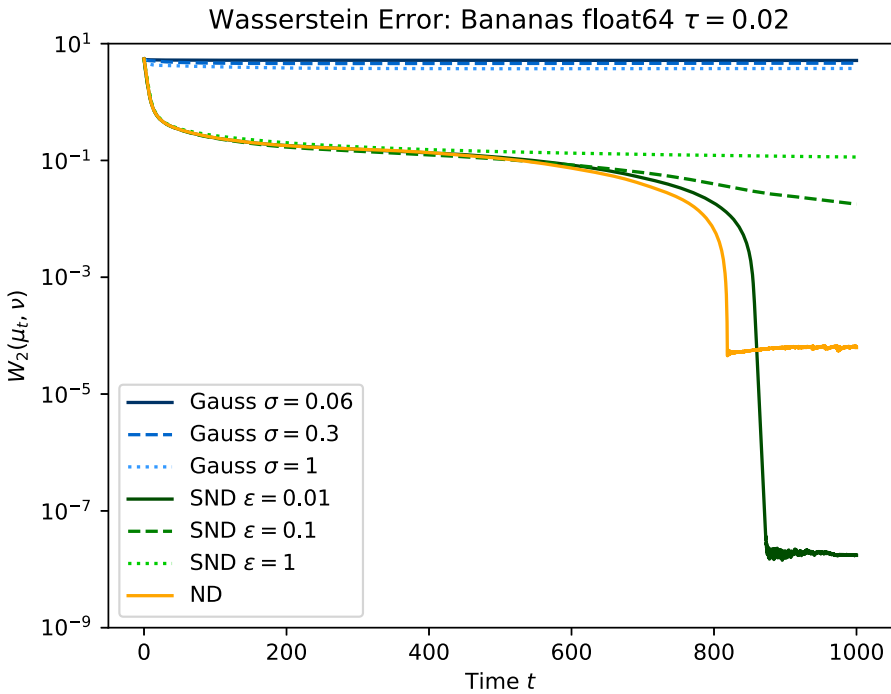


Fig. 6 W_2 error between Bananas target ν and flow γ_k^τ after k iterations with time $t = \tau k$. Computation with double precision with step size $\tau = 0.02$ shows a similar behavior as in Fig. 4

where $(\xi_p)_{p=1}^P \in (\mathbb{S}^{d-1})^P$ are equidistributed quadrature nodes on \mathbb{S}^{d-1} . This is a collection of P one-dimensional kernel sums. Each of them can be computed efficiently in $O((N + M) \log(N + M))$ operations, e.g., via fast Fourier summation [43] or, if F is the ND kernel, just by sorting [26]. Hence, (29) is more efficient if P is considerably smaller than the number of points. More details on the slicing summation and error estimates are provided in [28]. Furthermore, for our SND kernel, the explicit expression for F in Proposition 4.7 is somewhat cumbersome for large dimension d , while the slicing summation (29) only requires to evaluate f' .

Setup The initialization $x_n^{(0)}$, $n = 1, \dots, N$, are iid samples from a uniform distribution on $[0, 1]^d$. We compute the MMD flows (26) with $2^{15} = 32768$ iterations for the SND kernel $F = -C_{784} \mathcal{I}_{784}[\text{abs} * M_{2,\varepsilon}]$ with $\varepsilon \in \{0.001, 0.01, 0.1\}$ and the ND kernel $F = -\text{abs}$. The step size is $\tau = 1$, and the computations are performed in single precision. We use slicing summation (29) with $P = 785$ directions ξ_p that are the vertices of the centrally symmetric simplex, to which we apply a random rotation in each iteration step (cf. [28]). The slicing summation requires only the sliced kernel

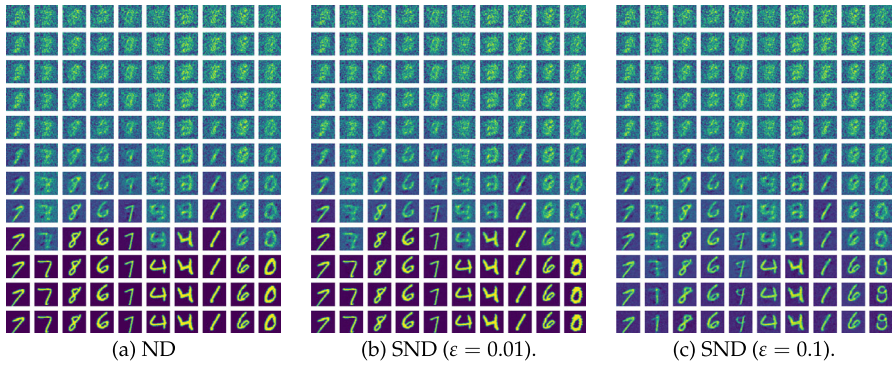


Fig. 7 MMD flow for MNIST target with different kernels. Each row shows the first 10 images $x_n \in \mathbb{R}^{28 \times 28}$, the ℓ -th row corresponds to the iteration $k = 2^{3+\ell}$, $\ell = 1, \dots, 12$

abs $*M_{2,\varepsilon}$ given in (7), but not the representation of F , which becomes quite clumsy in general (see Proposition 4.7).

The resulting images are shown in Fig. 7, where we see the MMDs for the SND with small ε and the Riesz kernel work comparably well and converge to the target measure ν . For a larger smoothing parameter ε , the flow needs more iterations to converge.

The distance to the target measured in the Wasserstein and MMD metrics is shown in Fig. 8. Here, we use the sliced approximation of the MMD. Note that the MMD, which is the objective we minimize, still depends on the kernel K . We see a similar behavior as for the previous low-dimensional examples with the error plateauing at some level, which becomes better for smaller ε , even slightly beating the ND kernel.

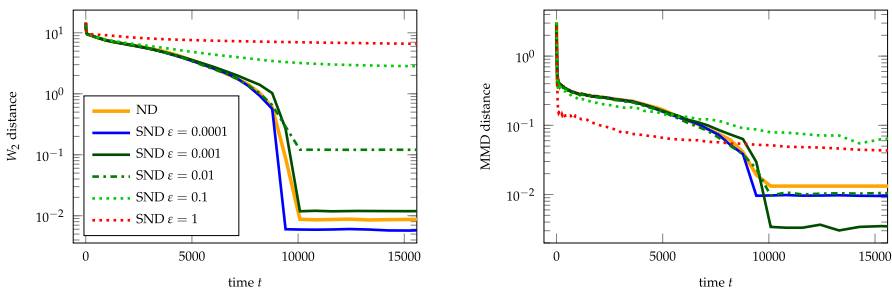


Fig. 8 MMD flow for MNIST target. Left: Wasserstein distance $W_2(\gamma^{(k)}, \nu)$. Right: MMD distance $\frac{1}{2} D_K^2(\gamma^{(k)}, \nu)$ for the respective kernels K

9 Conclusions

We introduced a smoothed negative distance kernel as an alternative to the negative distance kernel in MMDs. The novel kernel retains desired numerical properties of the negative distance kernel, but comes with well-defined gradient expressions and theoretical convergence guarantees of the corresponding gradient flow schemes. Therefore, our novel kernel appears to be well suited for various applications.

Concerning our future work, it may be interesting to examine if our kernel can also be used in Stein variational gradient descent [32, 42], where negative distance kernels do neither theoretically nor practically work.

Appendix 1. Fourier transform of tempered distributions

Let $\mathcal{S}'(\mathbb{R}^d)$ denote the space of tempered distributions, i.e., of linear functionals T on $\mathcal{S}(\mathbb{R}^d)$ fulfilling

$$\varphi_k \xrightarrow{\mathcal{S}} \varphi \implies \lim_{k \rightarrow \infty} \langle T, \varphi_k \rangle = \langle T, \varphi \rangle,$$

where $\xrightarrow{\mathcal{S}}$ denotes the convergence with respect to

$$\|\varphi\|_m := \max_{|\beta| \leq m} \|(1 + \|x\|_2)^m D^\beta \varphi(x)\|_{\mathcal{C}_0(\mathbb{R}^d)} \quad \text{for all } m \in \mathbb{N}.$$

In particular, $\mathcal{S}'(\mathbb{R}^d)$ contains all slowly increasing functions f , i.e., the functions fulfilling $|f(x)| \leq C(1 + \|x\|^N)$ for some $N \in \mathbb{N}$ and all functions in $L^p(\mathbb{R}^d)$, $p \in [1, \infty)$. As usual, for distributions of function type, the distribution T_f is identified with the function itself and the dual pairing becomes

$$\langle T_f, \varphi \rangle = \int_{\mathbb{R}^d} f \varphi \, dx \quad \text{for all } \varphi \in \mathcal{S}(\mathbb{R}^d).$$

The Fourier transform $\mathcal{F} : \mathcal{S}'(\mathbb{R}^d) \rightarrow \mathcal{S}'(\mathbb{R}^d)$, $T \mapsto \hat{T}$ is defined by

$$\langle T, \hat{\varphi} \rangle = \langle \hat{T}, \varphi \rangle \quad \text{for all } \varphi \in \mathcal{S}(\mathbb{R}^d).$$

In particular, we have for $f \in L^1(\mathbb{R}^d)$ in the above sense that $\hat{T}_f = T_{\hat{f}}$ with \hat{f} given by (1).

Example A.1 (Distributional versus generalized Fourier transform of polynomials) Let $p : \mathbb{R} \rightarrow \mathbb{R}$, $p(x) := \sum_{k=0}^{r-1} p_k x^k$. Then

$$\int_{\mathbb{R}^d} p(x) \hat{\varphi}(x) \, dx = 0 \quad \text{for all } \varphi \in \mathcal{S}_r,$$

so the generalized Fourier transform of p of order r is the zero function (see [57, Prop. 8.10]). In contrast, the distributional Fourier transform of p is given by

$$\hat{p} = \sum_{k=0}^{r-1} \left(\frac{i}{2\pi}\right)^k p_k \delta^{(k)}.$$

If we test only against functions in \mathcal{S}_r , both approaches coincide.

Example A.2 (Distributional Fourier transform of abs) Since abs is slowly increasing, it is a tempered distribution. Its distributional Fourier transform can be written as the distributional derivative of the Cauchy principal value,

$$\widehat{\text{abs}} = \frac{1}{2\pi^2} \left(\text{pv} \left(\frac{1}{\cdot}\right)\right)',$$

where

$$\left\langle \text{pv} \left(\frac{1}{\cdot}\right), \varphi \right\rangle := \lim_{\varepsilon \searrow 0} \int_{|x|>\varepsilon} \frac{\varphi(x)}{x} dx = \int_{\mathbb{R}} \frac{\varphi(x) - \varphi(0)}{x} dx, \quad \varphi \in \mathcal{S}(\mathbb{R}),$$

(see [43, Sect 4.3]) [19]. This can also be represented as the so-called Hadamard finite part $\text{H} \left(\frac{-1}{2\pi^2(\cdot)^2}\right)$ (see [13]), given by

$$\langle \widehat{\text{abs}}, \varphi \rangle = \left\langle \text{H} \left(\frac{-1}{2\pi^2(\cdot)^2}\right), \varphi \right\rangle := - \int_{\mathbb{R}} \frac{\varphi(\omega) - \varphi(0) - \varphi'(0)\omega}{2\pi^2\omega^2} d\omega.$$

If we test only against functions from $\varphi \in \mathcal{S}_2(\mathbb{R})$, we have $\varphi(0) = \varphi'(0) = 0$, so that this coincides with the generalized Fourier transform (3).

Another special case of tempered distributions is finite Borel measures $\mathcal{M}(\mathbb{R}^d)$ (see [43, Sect. 4.4]). More precisely, since $\mathcal{S}(\mathbb{R}^d)$ is a dense subspace of $(\mathcal{C}_0(\mathbb{R}^d), \|\cdot\|_\infty)$, we know by the Riesz representation theorem that $\mu \in \mathcal{M}(\mathbb{R}^d)$ can be identified with a tempered distribution $T_\mu : \mathcal{S}(\mathbb{R}^d) \rightarrow \mathbb{C}$ which acts on any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ by

$$\langle T_\mu, \varphi \rangle := \int_{\mathbb{R}^d} \varphi d\mu.$$

The Fourier transform on $\mathcal{M}(\mathbb{R}^d)$ is defined by $\mathcal{F} : \mathcal{M}(\mathbb{R}^d) \rightarrow \mathcal{C}_b(\mathbb{R}^d)$ with

$$\mathcal{F}\mu(\omega) = \hat{\mu}(\omega) := \int_{\mathbb{R}^d} e^{-2\pi i\omega \cdot} d\mu, \quad \omega \in \mathbb{R}^d, \tag{30}$$

and we have $\hat{T}_\mu = T_{\hat{\mu}}$. For positive measures $\mu \in \mathcal{M}(\mathbb{R}^d)$, i.e., $\mu(B) \geq 0$ for all Borel sets $B \subseteq \mathbb{R}^d$, we obtain a one-to-one mapping to positive definite functions by Bochner’s theorem.

Theorem A.3 (Bochner) *Any positive definite function $f: \mathbb{R}^d \rightarrow \mathbb{C}$ is the Fourier transform of a positive measure and conversely. If in addition $f(0) = 1$, then it is the Fourier transform of a probability measure.*

Note that, by our definition, positive definite functions automatically are continuous.

Appendix 2. Relation with Moreau envelopes

For a proper, convex, lower semi-continuous function $g: \mathbb{R}^d \rightarrow \mathbb{R}$ and $\lambda > 0$, the proximal function $\text{prox}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is defined by

$$\text{prox}_{\lambda g}(x) = \arg \min_{y \in \mathbb{R}^d} \left\{ \frac{1}{2} \|x - y\|^2 + \lambda g(y) \right\}$$

and its Moreau envelope by

$$H_{\lambda g}(x) = \min_{y \in \mathbb{R}^d} \left\{ \frac{1}{2} \|x - y\|^2 + \lambda g(y) \right\}.$$

The Moreau envelope is differentiable and

$$\nabla H_{\lambda g}(x) = x - \text{prox}_{\lambda g}(x),$$

so that

$$\text{prox}_{\lambda g}(x) = x - \nabla H_{\lambda g}(x) = \underbrace{\nabla \left(\frac{1}{2} \|x\|^2 - H_{\lambda g}(x) \right)}_{\psi(x)}. \tag{31}$$

Conversely, we have the following result of Moreau [38, Cor 10c].

Proposition B.1 *A function $G: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the proximal function of a proper, convex, lower semi-continuous function if and only if (i) there exists a convex differentiable function ψ such that $G = \nabla \psi$, and (ii) G is nonexpansive, i.e., $\|G(x) - G(y)\| \leq \|x - y\|$ for all $x, y \in \mathbb{R}^d$.*

In particular, we obtain for $g = \text{abs}$ that

$$\text{prox}_{\lambda \text{abs}}(x) = \begin{cases} x - \lambda & x > \lambda, \\ 0 & x \in [-\lambda, \lambda], \\ x + \lambda & x < -\lambda, \end{cases}$$

and the Moreau envelope

$$H_{\lambda \text{ abs}}(x) = \begin{cases} \frac{1}{2}x^2 & |x| \leq \lambda \\ \lambda(|x| - \frac{\lambda}{2}) & \text{otherwise} \end{cases}$$

is known as the *Huber function*.

On the other hand, we have

$$(\text{abs} * M_1)(x) = \begin{cases} x^2 + \frac{1}{4} & |x| \leq \frac{1}{2}, \\ |x| & |x| > \frac{1}{2}, \end{cases}$$

so that by Proposition 3.1, for $\varepsilon = 2\lambda$,

$$(\text{abs} * M_{1,2\lambda})(x) = \frac{1}{\lambda} H_{\lambda \text{ abs}}(x) + \frac{\lambda}{2} = \begin{cases} \frac{x^2}{2\lambda} + \frac{\lambda}{2} & |x| \leq \lambda, \\ |x| & |x| > \lambda. \end{cases}$$

Thus, by (3) and Propositions 2.2 and 3.2, the generalized Fourier transform of the Huber function is given by

$$\hat{H}_{\lambda \text{ abs}}(\omega) = -\frac{\lambda}{2\pi^2\omega^2} \text{sinc}(2\lambda\omega). \tag{32}$$

Since this function has positive and negative values, we conclude by Proposition 2.3 that the negative Huber function is not conditionally positive definite of any order.

Let us see if $\text{abs} * M_{1,\varepsilon}$ is the Moreau envelope of some function. Regarding (31), we consider

$$\psi(x) := \frac{1}{2}x^2 - (\text{abs} * M_{1,2\lambda})(x) = \begin{cases} \frac{1}{2}(1 - \frac{1}{\lambda})x^2 - \frac{\lambda}{2} & |x| \leq \lambda, \\ \frac{x^2}{2} - |x| & |x| > \lambda, \end{cases}$$

which is convex for $\lambda \geq 1$ and has a nonexpansive derivative

$$\psi'(x) = x - (\text{abs} * M_{1,\varepsilon})'(x) = \begin{cases} (1 - \frac{1}{\lambda})x & |x| \leq \lambda, \\ x - 1 & x > \lambda, \\ x + 1 & x < -\lambda. \end{cases}$$

Thus, by Moreau’s Proposition B.1, we see that $\text{abs} * M_{1,2\lambda}$ is a Moreau envelope if and only if $\varepsilon = 2\lambda \geq 2$. More generally, we have the following proposition.

Proposition B.2 *For $m \in \mathbb{N}$, $m \geq 1$, the function $\text{abs} * M_{m,\varepsilon}$ is the Moreau envelope of a proper, convex, lower semi-continuous function if and only if $\varepsilon \geq 2M_m(0)$.*

Proof By the above considerations, the assertion is true for $m = 1$. By Proposition 3.1, the function

$$\psi(x) := \frac{1}{2}x^2 - (\text{abs} * M_{m,\varepsilon})(x)$$

fulfills

$$\psi''(x) = 1 - \frac{2}{\varepsilon} M_m \left(\frac{x}{\varepsilon} \right) \geq 1 - \frac{2}{\varepsilon} M_m(0) \geq 0$$

if and only if $\varepsilon \geq 2M_m(0)$, and exactly in this case, ψ is convex. Further, because $\psi'' \leq 1$, we see that ψ' is nonexpansive and by Moreau’s Proposition B.1, and the function $\text{abs} * M_{m,\varepsilon}$ is a Moreau envelope. □

Appendix 3. Proofs

Proofs from Section 2

Proof of Proposition 2.2 Since $f \in \mathcal{C}(\mathbb{R}^d)$ is slowly increasing and $u \in \mathcal{C}_c(\mathbb{R}^d)$, we conclude by straightforward computations that $f * u$ is continuous and slowly increasing, too. Therefore, $\langle f * u, \hat{\varphi} \rangle$ exists for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Using Fubini’s theorem, we obtain

$$\begin{aligned} \langle f * u, \hat{\varphi} \rangle &= \int_{\mathbb{R}^d} (f * u)(x) \hat{\varphi}(x) \, dx = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} u(y) f(x - y) \, dy \hat{\varphi}(x) \, dx \\ &= \int_{\mathbb{R}^d} u(y) \int_{\mathbb{R}^d} f(x - y) \hat{\varphi}(x) \, dx \, dy. \end{aligned}$$

By the translation-modulation theorem, we know that

$$\hat{\varphi}(x) = \mathcal{F}[e^{-2\pi i \langle \cdot, y \rangle} \varphi](x - y),$$

so that

$$\begin{aligned} \langle f * u, \hat{\varphi} \rangle &= \int_{\mathbb{R}^d} u(y) \int_{\mathbb{R}^d} f(x - y) \mathcal{F}[e^{-2\pi i \langle \cdot, y \rangle} \varphi](x - y) \, dx \, dy \\ &= \int_{\mathbb{R}^d} u(y) \int_{\mathbb{R}^d} f(x) \mathcal{F}[e^{-2\pi i \langle \cdot, y \rangle} \varphi](x) \, dx \, dy. \end{aligned}$$

Since f has a generalized Fourier transform \hat{f} of order r , this implies for $\varphi \in \mathcal{S}_{2r}(\mathbb{R}^d)$ that

$$\begin{aligned} \langle f * u, \hat{\varphi} \rangle &= \int_{\mathbb{R}^d} \hat{f}(x) \varphi(x) \int_{\mathbb{R}^d} u(y) e^{-2\pi i \langle x, y \rangle} \, dy \, dx \\ &= \int_{\mathbb{R}^d} \hat{f}(x) \varphi(x) \hat{u}(x) \, dx. \end{aligned}$$

Hence, $f * u$ has a generalized Fourier transform of order r , namely $\hat{f} \hat{u}$. □

Proofs from Section 3

Proof of Proposition 3.1 i) follows directly by definition of f and since u is even.

To show ii), let $x > R := \text{diam}(\text{supp } u)/2$. The case $x < -R$ follows similarly. Then, we obtain

$$(\text{abs} * u)(x) = \int_{-R}^R u(y)(x - y) \, dy = x \int_{-R}^R u(y) \, dy - \int_{-R}^R y u(y) \, dy = x \cdot 1 - 0 = x.$$

In iii), we only have to show that $f'' = 2u$. Then, the smoothness of f follows by $u \in C^n(\mathbb{R})$. Using Lebesgue’s dominated convergence theorem, we conclude

$$\begin{aligned} \frac{d}{dx} (\text{abs} * u)(x) &= \lim_{h \rightarrow 0} \int_{-R}^R \underbrace{\frac{|x + h - y| - |x - y|}{h}}_{|g_{x,h}| \leq u} u(y) \, dy = \int_{-R}^R \text{sgn}(x - y) u(y) \, dy \\ &= (\text{sgn} * u)(x), \end{aligned}$$

where

$$\text{sgn}(x) := \begin{cases} 1 & x \geq 0, \\ -1 & x < 0. \end{cases}$$

The right derivative of sgn is given by

$$\lim_{h \searrow 0} \frac{\text{sgn}(x + h) - \text{sgn}(x)}{h} = \lim_{h \searrow 0} \frac{2}{h} \mathbb{1}_{[-h,0)}(x).$$

Therefore, we have by continuity of u that

$$\begin{aligned} \lim_{h \searrow 0} \frac{(\text{sgn} * u)(x + h) - (\text{sgn} * u)(x)}{h} &= \lim_{h \searrow 0} \int_{\mathbb{R}} \frac{2}{h} \mathbb{1}_{[-h,0)}(y) u(x - y) \, dy \\ &= 2 \lim_{h \searrow 0} \frac{1}{h} \int_{-h}^0 u(x - y) \, dy = 2u(x). \end{aligned}$$

We obtain the same result for the left derivative. Since $f'' = 2u \geq 0$, the function f is convex.

For iv), we have by Lemma 2.2 and (3) that

$$-\hat{f} = \mathcal{F}[-\text{abs} * u] = (-\widehat{\text{abs}}) \hat{u} \geq 0.$$

Therefore, $-f$ is conditionally positive definite of order $r = 1$ by Theorem 2.3.

Assertion v) follows by straightforward computation.

Finally, we show vi). Note that we cannot apply the usual convergence theorems for approximate identities in vi), because $\text{abs} \notin \mathcal{C}_0(\mathbb{R})$.

Let $\text{supp } u \subseteq [-R, R]$ for some $R > 0$. Then, we have by v) and ii) that

$$(\text{abs} * u_\varepsilon)(x) = |x| \quad \text{for } |x| \geq \varepsilon R.$$

For $|x| \leq \varepsilon R$, we conclude using $\int_{\mathbb{R}} u_\varepsilon(y) \, dy = 1$ that

$$\begin{aligned} \left| |x| - (\text{abs} * u_\varepsilon)(x) \right| &= \left| \int_{\mathbb{R}} |x| u_\varepsilon(x - y) \, dy - \int_{\mathbb{R}} |y| u_\varepsilon(x - y) \, dy \right| \\ &\leq \int_{\mathbb{R}} |x - y| u_\varepsilon(x - y) \, dy = \int_{-\varepsilon R}^{\varepsilon R} |y| u_\varepsilon(y) \, dy \\ &\leq \varepsilon R \int_{-\varepsilon R}^{\varepsilon R} u_\varepsilon(y) \, dy = \varepsilon R, \end{aligned}$$

which shows the uniform convergence. □

Proof of Corollary 3.3 By (6) and since $f'' = 2M_m$, we obtain

$$f(x) = \frac{2}{(m + 1)!} \sum_{k=0}^m (-1)^k \binom{m}{k} \left(x - k + \frac{m}{2}\right)_+^{m+1} + ax + b$$

with some $a, b \in \mathbb{R}$. Further, for $x \leq -\frac{m}{2}$, we conclude by $f(x) = -x$ that

$$f(x) = ax + b = -x,$$

so that $a = -1$ and $b = 0$. □

Proofs from Section 4

Proof of Lemma 4.1 The function $f := \text{abs} * M_2$ has the generalized Fourier transform of order 1 given by $\hat{f}(r) = -\frac{\text{sinc}^2(r)}{2\pi^2 r^2}$. We define

$$g(r) := -\frac{1}{2\pi} \frac{1}{r} \frac{d}{dr} \hat{f}(r).$$

For integrable functions, it was proven in [22, Thm. 1.1] that $g(\|\cdot\|)$ is the 3-dimensional Fourier transform of $f(\|\cdot\|)$. However, since f is not integrable, we use the generalized Fourier transform to argue that $\mathcal{F}_3[f(\|\cdot\|)] = g(\|\cdot\|)$: for an even test function $\psi \in \mathcal{S}_2(\mathbb{R})$, we apply [22, Thm. 1.1] to obtain

$$\mathcal{F}_3[\psi(\|\cdot\|)](\|x\|) = -\frac{1}{2\pi} \frac{1}{\|x\|} \hat{\psi}'(\|x\|),$$

and with the surface area $\omega_2 = 4\pi$, integration by parts gives

$$\begin{aligned} \int_{\mathbb{R}^3} g(\|x\|)\psi(\|x\|) \, dx &= \omega_2 \int_0^\infty g(r)\psi(r)r^2 \, dr = - \int_0^\infty 2\hat{f}'(r)\psi(r)r \, dr \\ &= - \left[2\hat{f}(r)\psi(r)r \right]_0^\infty + \int_0^\infty 2\hat{f}(r) \frac{d}{dr}(\psi(r)r) \, dr. \end{aligned}$$

Since $\psi \in \mathcal{S}_2(\mathbb{R})$, the first summand $[2\hat{f}(r)\psi(r)r]_0^\infty$ vanishes. The derivative of the odd function $\psi(r)r$ is even and still in $\mathcal{S}_2(\mathbb{R})$. Thus, we get by (2) that

$$\begin{aligned} \int_{\mathbb{R}^3} g(\|x\|)\psi(\|x\|) \, dx &= \int_{\mathbb{R}} \hat{f}(r) \frac{d}{dr}(\psi(r)r) \, dr = - \int_{\mathbb{R}} f(r)r \frac{d}{dr} \hat{\psi}(r) \, dr \\ &= 4\pi \int_0^\infty f(r) \frac{-1}{2\pi r} \frac{d}{dr} \hat{\psi}(r)r^2 \, dr = \int_{\mathbb{R}^3} f(\|x\|)\mathcal{F}_3[\psi(\|\cdot\|)](\|x\|) \, dx, \end{aligned}$$

i.e., $\mathcal{F}_3[f(\|\cdot\|)] = g(\|\cdot\|)$. Next, we show that for all test functions $\varphi \in \mathcal{S}_2(\mathbb{R}^3)$, it holds

$$\int_{\mathbb{R}^3} g(\|x\|)\varphi(x) \, dx = \int_{\mathbb{R}^3} f(\|x\|)\hat{\varphi}(x) \, dx.$$

For an arbitrary $\varphi \in \mathcal{S}(\mathbb{R}^d)$, define the radial test function

$$\text{Rad } \varphi(x) := \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} \varphi(\|x\|\xi) \, d\xi.$$

In [46, Thm. 4.2 i)], it was shown that Rad is a continuous projection of $\mathcal{S}(\mathbb{R}^d)$ to the space of radial Schwartz functions $\mathcal{S}_{\text{rad}}(\mathbb{R}^d)$. By the uniqueness of rotational invariant measures on the sphere (see [44, (2.3)]), we have

$$\text{Rad } \varphi(x) = \int_{SO(d)} f(Rx) \, d\mathcal{U}_{SO(d)}(R),$$

where $\mathcal{U}_{SO(d)}$ is the uniform measure on the set $SO(d)$ of $d \times d$ rotation matrices. Since the Fourier transform commutes with rotations, we have $\text{Rad}[\mathcal{F}\varphi] = \mathcal{F}[\text{Rad } \varphi]$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$, so the operators \mathcal{F} and Rad commute. It is easy to see that for $\varphi \in \mathcal{S}_2(\mathbb{R}^3)$, we also have $\text{Rad } \varphi \in \mathcal{S}_2(\mathbb{R}^3)$. The action of the test functions φ and $\text{Rad } \varphi$ on $g(\|\cdot\|)$ is the same, because $g(\|\cdot\|)$ is radial. Hence, we have for all $\varphi \in \mathcal{S}_2(\mathbb{R}^3)$ that

$$\begin{aligned} \int_{\mathbb{R}^3} g(\|x\|)\varphi(x) \, dx &= \langle g(\|\cdot\|), \varphi \rangle = \langle g(\|\cdot\|), \text{Rad } \varphi \rangle = \langle f(\|\cdot\|), \mathcal{F}[\text{Rad } \varphi] \rangle \\ &= \langle f(\|\cdot\|), \text{Rad } \hat{\varphi} \rangle = \langle f(\|\cdot\|), \hat{\varphi} \rangle = \int_{\mathbb{R}^3} f(\|x\|)\hat{\varphi}(x) \, dx. \end{aligned}$$

Consequently, $f(\|\cdot\|)$ has the generalized Fourier transform $g(\|\cdot\|)$ of order 1. Theorem 2.3 shows that $-f(\|\cdot\|) \notin \text{CP}_1(\mathbb{R}^3)$, because g changes its sign. Since

$g(\|\cdot\|)$ is the generalized Fourier transform of $f(\|\cdot\|)$ for all $r \geq 1$, we see that $-f(\|\cdot\|) \notin \text{CP}_r(\mathbb{R}^3)$ for all $r \geq 1$. Since $-f(\|\cdot\|) \notin \text{CP}_1(\mathbb{R}^3)$, we have by [57, Prop. 8.2] that $-f(\|\cdot\|) \notin \text{CP}_r(\mathbb{R}^d)$ for all $r \geq 0$ and all $d \geq 3$. \square

Proof of Proposition 4.3 Assume that $f \in \text{CP}_r(\mathbb{R})$, then $f(x) \in \mathcal{O}(|x|^{2r})$, by [34, Cor 2.3]. The function F is well-defined, because f is continuous and is slowly increasing. Let $x_1, \dots, x_N \in \mathbb{R}^d$ and $a_1, \dots, a_n \in \mathbb{R}$ such that

$$\sum_{j=1}^N a_j P(x_j) = 0 \tag{33}$$

for all polynomials P on \mathbb{R}^d of degree $< r$. In particular, any polynomial $p(t) = \sum_{k=0}^{r-1} c_k t^k$ on \mathbb{R} determines for an arbitrary fixed $\xi \in \mathbb{S}^{d-1}$, a polynomial on \mathbb{R}^d of degree $< r$ by

$$P_\xi(x) := p(\langle \xi, x \rangle) = \sum_{k=0}^{r-1} c_k \langle \xi, x \rangle^k = \sum_{k=0}^{r-1} c_k \left(\sum_{l=1}^d \xi_l x_l \right)^k.$$

By (33), we have

$$\sum_{j=1}^N a_j P_\xi(x_j) = \sum_{j=1}^N a_j p(\langle \xi, x_j \rangle) = 0.$$

Since f is conditionally positive definite of order r , we know that

$$0 \leq \sum_{j,k=1}^N a_j a_k f(|\langle \xi, x_j \rangle - \langle \xi, x_k \rangle|),$$

so that by Theorem 4.2 also

$$0 \leq \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} \sum_{j,k=1}^N a_j a_k f(|\langle \xi, x_j - x_k \rangle|) d\xi d\xi = \sum_{j,k=1}^N a_j a_k F(\|x_j - x_k\|).$$

Hence, $F(\|\cdot\|)$ is conditionally positive definite of order r . \square

Proof of Proposition 4.4 In [46, Eq. (6) & (7)], two operators were introduced: the rotation operator \mathcal{R}_d acts on a function $F: [0, \infty) \rightarrow \mathbb{R}$ as $\mathcal{R}_d F(x) := F(\|x\|)$, and the spherical averaging operator \mathcal{A}_d assigns to a function $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}$ integrable on the spheres $t \mathbb{S}^{d-1}$ for all $t > 0$ the function

$$\mathcal{A}_d \Phi(t) := \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} \Phi(t\xi) d\xi.$$

For $d = 1$, the spherical averaging operator reduces to $\mathcal{A}_1 \Phi(t) = \frac{1}{2}(\Phi(t) + \Phi(-t))$. Moving to distributions, the operator \mathcal{R}_d^* acts on a tempered distribution T as

$$\langle \mathcal{R}_d^* T, \psi \rangle = \langle T, (\mathcal{R}_d \circ \mathcal{A}_1) \psi \rangle \quad \text{for all } \psi \in \mathcal{S}(\mathbb{R}).$$

Since $F(\|\cdot\|)$ is continuous and slowly increasing, it can be identified with a tempered distribution. Let \mathcal{F}_d denote the Fourier transform of tempered distributions. Since F is $\lfloor \frac{d}{2} \rfloor$ -times continuously differentiable, we have by [46, Cor. 4.9] that

$$f := (\mathcal{F}_1 \circ \mathcal{R}_d^* \circ \mathcal{F}_d^{-1})[\mathcal{R}_d F]$$

is a distribution arising from a continuous, even function which satisfies $\mathcal{I}_d[f] = F$.

Let $\psi \in \mathcal{S}_{2r}(\mathbb{R})$ be an even. Then, $(\mathcal{R}_d \circ \mathcal{A}_1)\psi = \psi(\|\cdot\|)$ is a radial Schwartz function in $\mathcal{S}_{2r}(\mathbb{R}^d)$. Since $\mathcal{R}_d F$ has a Generalized Fourier transform $\rho(\|\cdot\|)$ of order r , we obtain

$$\begin{aligned} \int_{\mathbb{R}} f(r)\hat{\psi}(r) \, dr &= \langle f, \hat{\psi} \rangle = \langle \hat{f}, \psi \rangle = \langle (\mathcal{R}_d^* \circ \mathcal{F}_d^{-1})[\mathcal{R}_d F], \psi \rangle = \langle \mathcal{R}_d F, (\mathcal{F}_d^{-1} \circ \mathcal{R}_d \circ \mathcal{A}_1)\psi \rangle \\ &= \int_{\mathbb{R}^d} F(\|x\|)\mathcal{F}_d[\psi(\|\cdot\|)](x) \, dx = \int_{\mathbb{R}^d} \rho(\|x\|)\psi(\|x\|) \, dx \\ &= \frac{w_{d-1}}{2} \int_{\mathbb{R}} \rho(\omega)|\omega|^{d-1} \psi(\omega) \, d\omega. \end{aligned}$$

In particular, f has the generalized Fourier transform $\frac{w_{d-1}}{2} \rho(\omega)|\omega|^{d-1} \in \mathcal{C}(\mathbb{R} \setminus \{0\})$ of order r , which is nonnegative, so that f is conditionally positive definite of order r . □

Proof of Proposition 4.6 For $d \geq 2$, the term $(1 - t^2)^{\frac{d-3}{2}}$, $t \in [0, 1]$ is integrable. Since $f = \text{abs} * u_\varepsilon \in \mathcal{C}^n(\mathbb{R}) \subseteq L^\infty_{\text{loc}}(\mathbb{R})$, $n \in \mathbb{N}$, the function $F = \mathcal{I}_d[f]$ is well-defined. By Proposition 3.1, we know that f is nonnegative and even. Hence, F is also nonnegative and even. By Leibniz’s integral rule and since $f \in \mathcal{C}^{n+1}(\mathbb{R})$, we obtain for $k = 1, \dots, n + 2$ that

$$\begin{aligned} \frac{d^k}{ds^k} F(s) &= \frac{d^k}{ds^k} c_d \int_0^1 f(ts)(1 - t^2)^{\frac{d-3}{2}} \, dt = c_d \int_0^1 \frac{d^k}{ds^k} f(ts)(1 - t^2)^{\frac{d-3}{2}} \, dt \\ &= c_d \int_0^1 t^k f^{(k)}(ts)(1 - t^2)^{\frac{d-3}{2}} \, dt, \end{aligned}$$

so that $F \in \mathcal{C}^{n+2}(\mathbb{R})$. Since f is at least twice differentiable and $f(t) = \text{abs}(t)$ for $|t|$ large enough, it follows, that $\|f'\|_\infty < \infty$. For the first derivative of F we get

$$|F'(s)| = \left| c_d \int_0^1 t f'(ts)(1 - t^2)^{\frac{d-3}{2}} \, dt \right| \leq c_d \|(1 - t^2)^{\frac{d-3}{2}}\|_{L^1(0,1)} \|f'\|_\infty < \infty.$$

The convexity of F follows directly from the convexity of f .

Let $\text{supp}(u) \subseteq [-R, R]$ for some $R > 0$. Then, we have by Proposition 3.1 for $s \geq R$ that $f(s) = s$. Further, by Lemma 4.5, it holds $\mathcal{I}_d[\text{abs}] = C_d \text{abs}$. Hence, we obtain for $s > R$ that

$$\begin{aligned} |F(s) - C_d \text{abs}(s)| &= \left| c_d \int_0^1 (f(st) - \text{abs}(st))(1 - t^2)^{\frac{d-3}{2}} dt \right| \\ &\leq c_d \int_0^{\frac{R}{s}} |f(st) - st|(1 - t^2)^{\frac{d-3}{2}} dt \\ &= \frac{c_d}{s} \int_0^R |f(t) - t| \left(1 - \frac{t^2}{s^2}\right)^{\frac{d-3}{2}} dt \in \mathcal{O}\left(\frac{1}{s}\right). \end{aligned}$$

In particular, it holds $h := F - C_d \text{abs} \in \mathcal{C}_0(\mathbb{R}) \cap L^2(\mathbb{R})$.

Finally, we obtain by Proposition 3.1 that

$$\begin{aligned} F^\varepsilon(s) &= c_d \int_0^1 (\text{abs} * u_\varepsilon)(ts)(1 - t^2)^{\frac{d-3}{2}} dt \\ &= c_d \varepsilon \int_0^1 f\left(\frac{st}{\varepsilon}\right) (1 - t^2)^{\frac{d-3}{2}} dt = \varepsilon F\left(\frac{s}{\varepsilon}\right). \end{aligned}$$

and further

$$\begin{aligned} \|F^\varepsilon - C_d \text{abs}\|_{L^2(\mathbb{R})}^2 &\leq \int_{\mathbb{R}} |F^\varepsilon(s) - C_d \text{abs}(s)|^2 ds = \varepsilon^2 \int_{\mathbb{R}} \left|F\left(\frac{s}{\varepsilon}\right) - C_d \text{abs}\left(\frac{s}{\varepsilon}\right)\right|^2 ds \\ &= \varepsilon^2 \int_{\mathbb{R}} |h\left(\frac{s}{\varepsilon}\right)|^2 ds = \varepsilon^3 \|h\|_{L^2(\mathbb{R})}^2. \end{aligned}$$

This gives us the order of convergence in $L_2(\mathbb{R})$. The pointwise convergence of F^ε directly follows from (8). □

Proof of Proposition 4.7 By Corollary 3.3, we have

$$f(x) = \frac{2}{(m+1)!} \sum_{k=0}^m (-1)^k \binom{m}{k} \left(x - k + \frac{m}{2}\right)_+^{m+1} - x, \quad x \in \mathbb{R}.$$

Defining for $m \in \mathbb{N}$ and $a \in \mathbb{R}$ the function

$$b_{m,a}(x) = (x - a)_+^m,$$

we have

$$f(x) = \frac{2}{(m+1)!} \sum_{k=0}^m (-1)^k \binom{m}{k} b_{m+1, k - \frac{m}{2}}(x) - x.$$

If $a \leq 0$, we have $b_{m,a}(x) = (x - a)^m$ for $x \geq 0$. Then,

$$\begin{aligned} \mathcal{I}_d[b_{m,k}](s) &= c_d \int_0^1 f_{m,a}(st)(1 - t^2)^{\frac{d-3}{2}} dt = c_d \int_0^1 (st - a)^m(1 - t^2)^{\frac{d-3}{2}} dt \\ &= c_d \sum_{n=0}^m \binom{m}{n} s^n (-a)^{m-n} \int_0^1 t^n (1 - t^2)^{\frac{d-3}{2}} dt \\ &= \frac{c_d}{2} \sum_{n=0}^m \binom{m}{n} s^n (-a)^{m-n} \int_0^1 t^{\frac{n-1}{2}} (1 - t)^{\frac{d-3}{2}} dt. \end{aligned}$$

Since the Beta function satisfies $B(a, b) = \int_0^1 t^{a-1}(1 - t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ (see [1]), we obtain

$$\mathcal{I}_d[b_{m,k}](s) = \frac{c_d}{2} \sum_{n=0}^m \binom{m}{n} s^n (-a)^{m-n} \frac{\Gamma(\frac{d-1}{2})\Gamma(\frac{n+1}{2})}{\Gamma(\frac{d+n}{2})}.$$

If $a > 0$ and $s \leq a$, we have $\mathcal{I}_d[f_{m,a}](s) = 0$. Otherwise, i.e., for $0 < a < s$, we have

$$\begin{aligned} \mathcal{I}_d[b_{m,a}](s) &= c_d \int_{a/s}^1 (st - a)^m(1 - t^2)^{\frac{d-3}{2}} dt \\ &= \frac{c_d}{2} \sum_{n=0}^m \binom{m}{n} s^n (-a)^{m-n} \int_{a^2/s^2}^1 t^{\frac{n-1}{2}} (1 - t)^{\frac{d-3}{2}} dt \\ &= \frac{c_d}{2} \sum_{n=0}^m \binom{m}{n} s^n (-a)^{m-n} (B(\frac{n+1}{2}, \frac{d-1}{2}) - B_{a^2/s^2}(\frac{n+1}{2}, \frac{d-1}{2})). \end{aligned}$$

The claim follows by collecting the terms and Lemma 4.5. □

Proof of Theorem 4.9

- i) Since $f \in \text{CP}_1(\mathbb{R})$ by Proposition 3.1, we obtain by Proposition 4.3 that $\Phi \in \text{CP}_1(\mathbb{R}^d)$.
- ii) By (9), we know that

$$\Phi(x) = F(\|x\|) = \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} f(\langle x, \xi \rangle) d\xi \quad \text{for all } x \in \mathbb{R}^d.$$

Hence, we get $\Phi(0) = F(0) = f(0) = -(\text{abs} * M_2)(0) < 0$.

- iii) By Proposition 4.6, we directly conclude iii).
- iv) Since f is $n + 2$ times continuously differentiable by Proposition 3.1 iii), we obtain for any multi-index $\alpha \in \mathbb{N}^d$ with $|\alpha| \leq n + 2$ that

$$\partial^\alpha \Phi(x) = \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} \xi^\alpha f^{|\alpha|}(\langle x, \xi \rangle) d\xi \quad \text{for all } x \in \mathbb{R}^d.$$

v) Since $f'' = 2u$ is bounded, the first derivative f' is $2\|u\|_\infty$ -Lipschitz continuous. Hence, for $x, y \in \mathbb{R}^d$, we can estimate by (iv))

$$\begin{aligned} |\partial_i \Phi(x) - \partial_i \Phi(y)| &\leq \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} |\xi_i| |f'(\langle x, \xi \rangle) - f'(\langle y, \xi \rangle)| \, d\xi \\ &\leq \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} 2\|u\|_\infty \|x - y\| \, d\xi = 2\|u\|_\infty \|x - y\|, \end{aligned}$$

so that we obtain

$$\|\nabla \Phi(x) - \nabla \Phi(y)\|^2 = \sum_{i=1}^d |\partial_i \Phi(x) - \partial_i \Phi(y)|^2 \leq d(2\|u\|_\infty \|x - y\|)^2.$$

vi) Since $\nabla \Phi$ is L -Lipschitz continuous with $L = \sqrt{d}2\|u\|_\infty$, we know by [40, Thm. 2.1.5] that Φ is $-L$ -convex. Furthermore, Φ is concave because, for $t \in [0, 1]$ and $x, y \in \mathbb{R}^d$, it holds by the concavity of f from Proposition 3.1 iii) that

$$\begin{aligned} \Phi((1-t)x + ty) &\geq \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} (1-t)f(\langle x, \xi \rangle) + tf(\langle y, \xi \rangle) \, d\xi \\ &= (1-t)\Phi(x) + t\Phi(y). \end{aligned}$$

□

Proofs of Section 6

Proof of Proposition 6.1 Recall that $f = -\text{abs} * u$ is even and $F = \mathcal{I}_d[f]$ satisfies (9). Let $\mu \in \mathcal{M}_{1/2}(\mathbb{R}^d)$, then the following integral exists

$$\begin{aligned} \|\mu\|_{\mathcal{H}_K}^2 &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(x, y) \, d\mu(x) \, d\mu(y) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (F(\|x - y\|) - F(\|x\|) - F(\|y\|)) \, d\mu(x) \, d\mu(y) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} f(\langle x - y, \xi \rangle) - f(\langle x, \xi \rangle) - f(\langle -y, \xi \rangle) + f(0) \, d\xi \, d\mu(x) \, d\mu(y) \\ &\quad - F(0)|\mu(\mathbb{R}^d)|^2. \end{aligned}$$

Denote by \mathcal{T}_y the translation operator $\mathcal{T}_y[g](y) = g(x - y)$, by \mathcal{M}_y the modulation operator $\mathcal{M}_y[g](x) = e^{-2\pi i \langle x, y \rangle} g(x)$ and by $g_m(x) = \sqrt{m/\pi} e^{-mx^2}$ the Gaussian approximate identity as in [57, Thm. 5.20]. Since f is continuous and slowly increasing, [57, Thm. 5.20 (4)] yields

$$f(\langle x - y, \xi \rangle) - f(\langle x, \xi \rangle) - f(\langle -y, \xi \rangle) + f(0) = \lim_{m \rightarrow \infty} (\text{id} - \mathcal{T}_{(\xi, x)})(\text{id} - \mathcal{T}_{(\xi, -y)})(g_m), f).$$

Let $\varphi_m := (\text{id} - \mathcal{M}_{-(\xi, x)})(\text{id} - \mathcal{M}_{-(\xi, -y)})[\hat{g}_m] \in \mathcal{S}_2(\mathbb{R}^1)$. The function f has the generalized Fourier transform

$$\hat{f}(r) = \frac{\hat{u}(r)}{2\pi^2 r^2}$$

of order 1 by Lemma 2.2 and (3). As g_m is even, we have $\hat{g}_m = g_m$ and for all $m \in \mathbb{N}$, $m \geq 1$ we can write

$$\begin{aligned} \langle (\text{id} - \mathcal{T}_{(\xi, x)})(\text{id} - \mathcal{T}_{(\xi, -y)})[g_m], f \rangle &= \langle \hat{\varphi}_m, f \rangle = \langle \varphi_m, \hat{f} \rangle \\ &= \int_{\mathbb{R}} (1 - e^{2\pi i(\xi, x)r})(1 - e^{-2\pi i(\xi, y)r})\hat{g}_m(r)\hat{f}(r) \, dr. \end{aligned}$$

Since u is continuous with compact support, its Fourier transform is bounded. The term $r \mapsto (1 - e^{2\pi i(\xi, x)r})(1 - e^{-2\pi i(\xi, y)r})$ is bounded and has a zero of order 2 at zero, so that

$$\hat{f}(r)(1 - e^{2\pi i(\xi, x)r})(1 - e^{-2\pi i(\xi, y)r}) = \frac{\hat{u}(r)}{2\pi^2 r^2}(1 - e^{2\pi i(\xi, x)r})(1 - e^{-2\pi i(\xi, y)r})$$

is integrable. Moreover, we have

$$|\hat{g}_m(r)| = e^{-\frac{\pi^2 r^2}{m}} \leq 1 \quad \text{for all } r \in \mathbb{R} \text{ and } m \in \mathbb{N}, m \geq 1.$$

Since \hat{g}_m converges pointwise to the constant 1, Lebesgue’s convergence theorem yields

$$f((x - y, \xi)) - f((x, \xi)) - f((-y, \xi)) + f(0) = \int_{\mathbb{R}} (1 - e^{2\pi i(\xi, x)r})(1 - e^{-2\pi i(\xi, y)r})\hat{f}(r) \, dr.$$

Further, we obtain using Fubini’s theorem

$$\begin{aligned} &\omega_{d-1}(\|\mu\|_{\mathcal{H}_K}^2 + F(0)\mu(\mathbb{R}^d)^2) \\ &= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \int_{\mathbb{R}^d} (1 - e^{2\pi i(r\xi, x)}) \, d\mu(x) \int_{\mathbb{R}^d} (1 - e^{-2\pi i(r\xi, y)}) \, d\mu(y) \hat{f}(r) \, dr \, d\xi \\ &= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} |\mu(\mathbb{R}^d) - \hat{\mu}(r\xi)|^2 \hat{f}(r) \, dr \, d\xi \\ &= 2 \int_{\mathbb{S}^{d-1}} \int_0^\infty |\mu(\mathbb{R}^d) - \hat{\mu}(r\xi)|^2 \frac{\hat{f}(\|r\xi\|)}{\|r\xi\|^{d-1}} r^{d-1} \, dr \, d\xi \\ &= 2 \int_{\mathbb{R}^d} |\mu(\mathbb{R}^d) - \hat{\mu}(x)|^2 \frac{\hat{f}(\|x\|)}{\|x\|^{d-1}} \, dx. \end{aligned}$$

Inserting \hat{f} , we can write

$$\|\mu\|_{\mathcal{H}_K}^2 = \frac{1}{\pi^2 \omega_{d-1}} \int_{\mathbb{R}^d} |\hat{\mu}(0) - \hat{\mu}(x)|^2 \frac{\hat{u}(\|x\|)}{\|x\|^{d+1}} dx - F(0)\hat{\mu}(0)^2. \tag{34}$$

Now, assume that $\mu \in \mathcal{M}_{1/2}(\mathbb{R}^d)$ with $\|\mu\|_{\mathcal{H}_K} = 0$. Because $F(0) < 0$ by Theorem 4.9 and $\hat{u} \geq 0$ by (4), both summands in (34) are nonnegative and therefore must vanish. The second term yields that $\mu(\mathbb{R}^d) = \hat{\mu} = 0$. Since $\text{supp } \hat{u}(\|\cdot\|) \cdot \|\cdot\|^{-(d+1)} = \mathbb{R}^d$ (cf. [43, Lem 2.39]), it follows that $\hat{\mu}$ is constant with $\hat{\mu} \equiv \hat{\mu}(0) = 0$. This implies $\mu = 0$ as the Fourier transform $\mathcal{F}: \mathcal{M}(\mathbb{R}^d) \rightarrow \mathcal{C}_b(\mathbb{R}^d)$ is injective. Consequently, the KME is injective, which means that K is characteristic. \square

Proof of Theorem 6.2 Since $\Phi(x) \in \mathcal{O}(\|x\|^\alpha)$, we can estimate $|\Phi(x)| \leq C(1 + \|x\|^\alpha)$ for all $x \in \mathbb{R}^d$. For $\alpha \geq 1$, we have by convexity of $\|\cdot\|^\alpha$ that

$$\|x + y\|^\alpha \leq 2^{\alpha-1}(\|x\|^\alpha + \|y\|^\alpha).$$

For $\alpha \in (0, 1)$, we define the function $f: [0, \infty) \rightarrow \mathbb{R}$ by $f(x) := x^\alpha$, which is concave and monotone increasing. Then, we obtain for $x, y \geq 0$ with $x + y > 0$ that

$$\begin{aligned} f(x) &\geq \frac{y}{x+y} f(0) + \frac{x}{x+y} f(x+y), \\ f(y) &\geq \frac{x}{x+y} f(0) + \frac{y}{x+y} f(x+y). \end{aligned}$$

Adding both equations yields

$$f(x) + f(y) \geq f(x+y).$$

Since f is monotone increasing, we obtain by the triangle inequality

$$\|x + y\|^\alpha \leq (\|x\| + \|y\|)^\alpha \leq \|x\|^\alpha + \|y\|^\alpha.$$

Summarizing, we have for $\alpha \geq 0$ that

$$\|x + y\|^\alpha \leq 2^\alpha(\|x\|^\alpha + \|y\|^\alpha).$$

Therefore, we can guarantee the existence of the integral

$$\begin{aligned} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\Phi(x-y)| d\sigma(x) d\sigma(y) &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} C(1 + 2^\alpha(\|x\|^\alpha + \|y\|^\alpha)) d\sigma(x) d\sigma(y) \\ &\leq C\sigma(\mathbb{R}^d) \left(\sigma(\mathbb{R}^d) + 2^{\alpha+1} \int_{\mathbb{R}^d} \|x\|^\alpha d\sigma(x) \right) < \infty. \end{aligned}$$

Hence, the discrepancy $d_{\tilde{\mathcal{K}}}(\mu, \nu)$ is well-defined for $\mu, \nu \in \mathcal{M}_\alpha(\mathbb{R}^d)$.

By (11), we see that $K(x, x) \in \mathcal{O}(\|x\|^{2\beta})$, $\beta := \max\{r - 1, (\alpha + r - 1)/2\}$ such that d_K is by (16) well-defined for measures in \mathcal{M}_β .

Now, assume additionally that the first $r - 1$ moments of μ and ν coincide. This implies that for all $p_j \in \Pi_{r-1}(\mathbb{R}^d)$ that

$$\int_{\mathbb{R}^d} p_j(x) \, d(\mu - \nu)(x) = 0.$$

Then, we obtain

$$\begin{aligned} d_K(\mu, \nu)^2 &= d_{\tilde{K}}(\mu, \nu)^2 - \sum_{j=1}^N \int_{\mathbb{R}^d} p_j(x) \, d(\mu - \nu)(x) \int_{\mathbb{R}^d} \Phi(\xi_j - y) \, d(\mu - \nu)(y) \\ &\quad - \sum_{k=1}^N \int_{\mathbb{R}^d} p_k(y) \, d(\mu - \nu)(y) \int_{\mathbb{R}^d} \Phi(x - \xi_k) \, d(\mu - \nu)(x) \\ &\quad + \sum_{k,j=1}^N \Phi(\xi_j - \xi_k) \int_{\mathbb{R}^d} p_j(x) \, d(\mu - \nu)(x) \int_{\mathbb{R}^d} p_k(y) \, d(\mu - \nu)(y) \\ &= d_{\tilde{K}}(\mu, \nu)^2. \end{aligned}$$

□

Proof of Proposition 6.4 1. First, we show that K is $\lfloor \frac{n+2}{2} \rfloor$ times continuously differentiable. Let $\alpha \in \mathbb{N}^d$ with $|\alpha| \leq \lfloor \frac{n+2}{2} \rfloor$. The case $|\alpha| = 0$ is clear. For $|\alpha| \geq 1$, we obtain

$$\begin{aligned} \partial_x^\alpha \partial_y^\alpha K(x, y) &= \partial_x^\alpha \partial_y^\alpha F(\|x - y\|) = \partial_x^\alpha \partial_y^\alpha \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} f(\langle \xi, x - y \rangle) \, d\xi \\ &= \frac{(-1)^{|\alpha|}}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} \xi^{2\alpha} f^{2|\alpha|}(\langle \xi, x - y \rangle) \, d\xi. \end{aligned} \tag{35}$$

By [52, Cor. 4.36], this implies that every $h \in \mathcal{H}_K$ is at least $\lfloor \frac{n+2}{2} \rfloor$ -times continuously differentiable.

2. For the second part, assume that $n \geq 2$. By [52, Lem 4.34] and (35), we obtain

$$\langle \partial_{x_i} K(x, \cdot), \partial_{x_i} K(y, \cdot) \rangle_{\mathcal{H}_K} = \partial_{x_i} \partial_{y_i} K(x, y) = \frac{-1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} \xi_i^2 f''(\langle \xi, x - y \rangle) \, d\xi.$$

By Proposition 3.1, the function f is even and $f''' = 2u'$. Hence, u' is odd and $\|u''\|_\infty$ -Lipschitz continuous and $f'''(0) = 2u'(0) = 0$. Thus, we obtain for $s > 0$ that

$$|f''(s) - f''(0)| \leq \int_0^s |f'''(t)| \, dt = \int_0^s |f'''(t) - f'''(0)| \, dt \leq \int_0^s 2\|u''\|_\infty t \, dt = \|u''\|_\infty s^2.$$

Hence, we can estimate

$$\begin{aligned} \|\partial_{x_i} K(x, \cdot) - \partial_{x_i} K(y, \cdot)\|_{\mathcal{H}_K}^2 &= \|\partial_{x_i} K(x, \cdot)\|_{\mathcal{H}_K}^2 + \|\partial_i K(y, \cdot)\|_{\mathcal{H}_K}^2 - 2\langle \partial_{x_i} K(x, \cdot), \partial_{x_i} K(y, \cdot) \rangle_{\mathcal{H}_K} \\ &= -\frac{2}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} \xi_i^2 (f''(0) - f''(\langle \xi, x - y \rangle)) \, d\xi \\ &\leq \frac{2\|u''\|_\infty}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} |\langle \xi, x - y \rangle|^2 \, d\xi \\ &\leq 2\|u''\|_\infty \|x - y\|^2. \end{aligned}$$

Therefore, $\partial_{x_i} K(x, \cdot)$ is Lipschitz continuous with constant $L := \sqrt{2\|u''\|_\infty}$. Finally, we see again by (the proof of) [52, Cor. 4.36], for any $h \in \mathcal{H}_K$, that

$$\begin{aligned} |\partial_{x_i} h(x) - \partial_{x_i} h(y)| &= |\langle h, \partial_{x_i} K(x, \cdot) \rangle_{\mathcal{H}_K} - \langle h, \partial_{x_i} K(y, \cdot) \rangle_{\mathcal{H}_K}| \\ &= |\langle h, \partial_{x_i} K(x, \cdot) - \partial_{x_i} K(y, \cdot) \rangle_{\mathcal{H}_K}| \\ &\leq \|h\|_{\mathcal{H}_K} L \|x - y\|, \end{aligned}$$

which gives the assertion by

$$\|\nabla h(x) - \nabla h(y)\| \leq \sqrt{d}L \|h\|_{\mathcal{H}_K} \|x - y\|.$$

□

Appendix 4. Geodesic λ -convexity of MMD functional with smoothed distance kernel

A *generalized geodesic* is an interpolating curve $\gamma_t: [0, 1] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ that connects two measures μ^2 and $\mu^3 \in \mathcal{P}_2(\mathbb{R}^d)$ via a three-plan μ . More specifically, for a base $\mu^1 \in \mathcal{P}(\mathbb{R}^d)$, this three-plan $\mu \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ has marginals $\pi_{\#}^i \mu = \mu^i, i = 1, 2, 3$ and must be optimal in the sense that $\pi_{\#}^{1,i} \mu \in \Pi_{\text{opt}}(\mu^1, \mu^i)$ for $i = 1, 2$. A generalized geodesic γ_t joining μ^2 with μ^3 via μ^1 is defined as $\gamma_t := ((1 - t)\pi^2 + t\pi^3)_{\#} \mu$. For any choice of $\mu^1, \mu^2, \mu^3 \in \mathcal{P}_2(\mathbb{R}^d)$, we can always find optimal plans $\mu^{1,i} \in \Pi_{\text{opt}}(\mu^1, \mu^i)$ for $i = 1, 2$, and by the Gluing Lemma [56], there exists a three plan $\mu \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ with marginals $\pi_{\#}^i \mu = \mu^i, i = 1, 2, 3$ and $\pi_{\#}^{1,i} \mu \in \Pi_{\text{opt}}(\mu^1, \mu^i)$ for $i = 1, 2$. This means that there always exists at least one generalized geodesic joining μ^2 with μ^3 via μ^1 . However, this generalized geodesic is not necessarily unique.

Given $\lambda \in \mathbb{R}$, a function $G: \mathcal{P}_2(\mathbb{R}^d) \rightarrow [0, \infty]$ is λ -convex along generalized geodesics, if for any choice $\mu^1, \mu^2, \mu^3 \in \text{dom}(G)$, there always exists a generalized geodesic γ_t joining μ^2 with μ^3 via μ^1 , such that

$$G(\gamma_t) \leq (1 - t)G(\mu^2) + tG(\mu^3) - \frac{\lambda}{2}t(1 - t) \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \|x_2 - x_3\|^2 \, d\mu(x_1, x_2, x_3).$$

For a more detailed description, we refer to [2, Section 9.2].

In [2], sufficient conditions for the λ -convexity of the following two typical energy functionals were given. The *potential energy* $V : \mathbb{R}^d \rightarrow \mathbb{R}$ is defined by

$$\mathcal{V}(\mu) := \int_{\mathbb{R}^d} V(x) \, d\mu(x).$$

Lemma D.1 *Let V be lower semi-continuous and have quadratic grow, i.e.,*

$$V(x) \geq -A - B\|x\|^2 \text{ for all } x \in \mathbb{R}^d$$

with $A, B \in \mathbb{R}$. If V is λ -convex, then \mathcal{V} is λ -convex along generalized geodesics.

For $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, the *interaction energy* is given by

$$\mathcal{K}(\mu) := \int_{\mathbb{R}^d \times \mathbb{R}^d} K(x, y) \, d\mu(x) \, d\mu(y).$$

Since the interaction energy can be seen as a potential energy on the product space, Proposition D.1 also applies to the interaction energy.

Lemma D.2 *Let K be lower semi-continuous and have quadratic grow, i.e.,*

$$K(x, y) \geq -A - B(\|x\|^2 + \|y\|^2) \text{ for all } x, y \in \mathbb{R}^d$$

with $A, B \in \mathbb{R}$. If K is λ -convex, then \mathcal{K} is λ -convex along generalized geodesics.

Let F be defined as in (10) and $K(x, y) = F(\|x - y\|)$. Then, the MMD functional G from (23) can be rewritten as

$$G(\mu) = \frac{1}{2}\mathcal{K}(\mu) + \mathcal{V}(\mu) + \frac{1}{2}c_v, \quad V(x) := - \int_{\mathbb{R}^d} F(\|x - y\|) \, d\nu(y) \quad (36)$$

where $c_v \geq 0$ is a constant. Both V and K suffice the conditions in Lemmas D.1 and D.2.

Proposition D.3 *Let F be defined as in (10) and V and K in (36), then \mathcal{V} and \mathcal{K} are lower semi-continuous and have quadratic grow. Moreover, V is convex and K is λ -convex with $\lambda = -4\sqrt{d}\|u\|_\infty$. In summary, the MMD functional G from (36) is lower semi-continuous and λ -convex with λ above.*

Proof By Theorem 4.9, we can write $F(s) = -C_d|s| + \varphi(s)$ with $\varphi \in \mathcal{C}_0(\mathbb{R})$. For the lower semi-continuity of V , we have

$$\begin{aligned} |V(x_1) - V(x_2)| &\leq \int_{\mathbb{R}^d} |F(\|x_1 - y\|) - F(\|x_2 - y\|)| \, d\nu(y) \\ &= \int_{\mathbb{R}^d} |C_d\|x_1 - y\| + \varphi(\|x_1 - y\|) - C_d\|x_2 - y\| - \varphi(\|x_2 - y\|)| \, d\nu(y) \end{aligned}$$

$$\begin{aligned} &\leq \int_{\mathbb{R}^d} C_d \|x_1 - y\| - \|x_2 - y\| + |\varphi(\|x_1 - y\|) - \varphi(\|x_2 - y\|)| \, d\nu(y) \\ &\leq C_d \|x_1 - x_2\| + \int_{\mathbb{R}^d} \|\varphi(\|x_1 - y\|) - \varphi(\|x_2 - y\|)\| \, d\nu(y). \end{aligned}$$

Since $\varphi \in C_0(\mathbb{R})$, it follows by Lebesgue’s dominated convergence theorem that V is continuous. Moreover, choosing $A, B = 0$, we see that V has quadratic grow. By Theorem 4.9 iii), the function $-F(\|x\|)$ is convex. For $x_1, x_2 \in \mathbb{R}^d$ and $t \in [0, 1]$, it holds

$$\begin{aligned} V((1-t)x_1 + tx_2) &= \int_{\mathbb{R}^d} -F(\|(1-t)(x_1 - y) + t(x_2 - y)\|) \, d\nu(y) \\ &\leq \int_{\mathbb{R}^d} -(1-t)F(\|x_1 - y\|) - tF(\|x_2 - y\|) \, d\nu(y) \\ &= (1-t)V(x_1) + tV(x_2). \end{aligned}$$

Hence V is convex, too.

For the interaction energy, it is clear that K is continuous, because F is continuous, and we can choose $A = \|\varphi\|_\infty + 2C_d$ and $B = C_d$ to obtain

$$\begin{aligned} F(\|x - y\|) &= -C_d \|x - y\| + \varphi(\|x - y\|) \geq -\|\varphi\|_\infty - C_d(\|x\| + \|y\|) \\ &\geq -\|\varphi\|_\infty - C_d(1 + \|x\|^2 + 1 + \|y\|^2) \geq -A - B(\|x\|^2 + \|y\|^2). \end{aligned}$$

By Corollary 4.9 v), we get that K is λ -convex with $\lambda = -4\sqrt{d}\|u\|_\infty$. □

By [2, 11.2.1b], a lower bounded λ -convex functional is always coercive, so that [2, Thm.11.2.1] can be formulated as follows:

Theorem D.4 *Let $G : \mathcal{P}_2(\mathbb{R}^d) \rightarrow [0, \infty]$ be proper lower semi-continuous and λ convex. Then, for $\gamma^{(0)} \in \text{dom}(G)$, there is a unique Wasserstein gradient flow γ_t starting in $\gamma^{(0)}$. Moreover, the piecewise constant curve $\gamma_t^\tau := \gamma^{(k)}$, $t \in ((k-1)\tau, k\tau]$ given by the implicit Euler scheme (JKO scheme)*

$$\gamma^{(k+1)} \in \arg \min_{\gamma \in \mathcal{P}_2(\mathbb{R}^d)} \frac{1}{2\tau} W_2^2(\gamma^{(k)}, \gamma) + \phi(\gamma),$$

converges locally uniformly to γ_t . In particular, this holds true for our MMD functional with smoothed kernel G in (36).

It was shown in [27, Prop. 9] that for certain functionals, e.g., G in (36), the so-called Wasserstein steepest descent flows (explicit scheme) and the Wasserstein gradient flows (implicit scheme) coincide.

Since the MMD functional with our SND kernel (same for the Gaussian kernel) is only λ -convex with $\lambda < 0$, it is not ensured that its Wasserstein gradient flow, resp. its approximation by an Euler forward scheme converges towards the target ν . Here is an example.

Example D.5 In general, it is not clear whether the gradient flow γ_t of the MMD functional with smooth kernels converges towards the target measure ν as $t \rightarrow \infty$. To this end, consider the symmetric setup with the target and initial measures

$$\begin{aligned} \nu &:= \frac{1}{2}(\delta_{y_1} + \delta_{y_2}), \quad y_1 = e_1, \quad y_2 = -e_1, \\ \mu &:= \frac{1}{2}(\delta_{x_1} + \delta_{x_2}), \quad x_1 = \frac{1}{\sqrt{3}}e_2, \quad x_2 = -\frac{1}{\sqrt{3}}e_2. \end{aligned}$$

Then, with $\tilde{F}(s) := \frac{F'(s)}{s}$, the velocity field becomes for $i = 1, 2$

$$\begin{aligned} v_t(x_i) &= \frac{1}{2}(x_i - x_j)\tilde{F}(\|x_i - x_j\|) - \frac{1}{2}\left((x_i - y_1)\tilde{F}(\|x_i - y_1\|) + (x_i - y_2)\tilde{F}(\|x_i - y_2\|)\right) \\ &= \frac{1}{\sqrt{3}}e_2\tilde{F}\left(\frac{2}{\sqrt{3}}\right) - \frac{1}{2}\left(\frac{2}{\sqrt{3}}e_2 - e_1 + e_1\right)\tilde{F}\left(\frac{2}{\sqrt{3}}\right) = 0, \end{aligned}$$

so that we get stuck in $\gamma_t = \mu$ for all $t \geq 0$. In general, ensuring convergence towards the target is challenging due to local extrema. However, in the numerical experiments, we observed that for both ND and SND, the flows typically performed well in approximating the target.

Proof of Proposition 7.5 The simplification of the iterations to

$$\begin{aligned} x^{(k+1)} &= x^{(k)} + \tau \frac{x^{(k)} - y}{\|x^{(k)} - y\|} F'(\|x^{(k)} - y\|) \tag{37} \\ &= y + (x^{(k)} - y) \left(1 + \tau \frac{F'(\|x^{(k)} - y\|)}{\|x^{(k)} - y\|}\right) \end{aligned}$$

is straightforward.

i) For $F = -\frac{1}{2} \text{abs}$, we have $F'(s) = -\frac{1}{2}$ for $s > 0$. Then, we obtain by (37) and since $\|x^{(0)} - y\| < \frac{\tau}{2}$ that

$$\|y - x^{(1)}\| = \frac{\tau}{2} - \|x^{(0)} - y\| < \frac{\tau}{2}.$$

The second step $x^{(2)}$ jumps exactly back to $x^{(0)}$ because

$$\begin{aligned} x^{(2)} &= y + (x^{(1)} - y) \left(1 - \frac{\tau}{2\|x^{(1)} - y\|}\right) \\ &= y + (x^{(0)} - y) \left(1 - \frac{\tau}{2\|x^{(0)} - y\|}\right) \left(1 - \frac{\tau}{2\|x^{(1)} - y\|}\right) \\ &= y + (x^{(0)} - y) \frac{(2\|x^{(0)} - y\| - \tau)(-2\|x^{(0)} - y\|)}{2\|x^{(0)} - y\|(\tau - 2\|x^{(0)} - y\|)} = x^{(0)}. \end{aligned}$$

Consequently, $(x^{(k)})_k$ oscillates between $x^{(0)}$ and $x^{(1)}$.

ii) Generally, for λ -convex functionals with $\lambda > 0$, Baillon-Haddad’s theorem [5, Cor. 18.17] ensures convergence of (27) for $\tau < \lambda^{-1}$. For completeness, we provide a simpler proof for our setting. Let $F = \mathcal{I}_d[-|u|]$, with $u \in \mathcal{U}^0(\mathbb{R})$. We know that F is convex and twice differentiable. In particular, we have for $\tilde{F}(s) := \frac{F'(s)}{s}$ that $\tilde{F}(0) = F''(0) < 0$. Since $\tilde{F} \in \mathcal{C}(\mathbb{R})$ by Proposition 4.6, we can find $\delta > 0$ such that $F(x) < \frac{1}{2}F''(0)$ for $|x| < \delta$. If we assume that $\|x^{(k)} - y\| < \tau$ and $\tau < \min\{\delta, \|\tilde{F}'\|_\infty^{-1}\}$, we obtain

$$\|x^{(k+1)} - y\| = \|x^{(k)} - y\| \left(1 + \tau \tilde{F}(\|x^{(k)} - y\|)\right).$$

We always have

$$1 + \tau \tilde{F}(\|x^{(k)} - y\|) > 1 - \tau \|\tilde{F}\|_\infty > 0.$$

Since $\|x^{(k)} - y\| < \delta$, we know that $F(\|x^{(k)} - y\|) < \frac{1}{2}F''(0) < 0$, which implies

$$1 + \tau \tilde{F}(\|x^{(k)} - y\|) < 1 + \frac{\tau}{2}F''(0) < 1.$$

This yields $\|x^{(k+1)} - y\| < \|x^{(k)} - y\| < \delta$, and thus, by induction,

$$\|x^{(k+1)} - y\| \leq \|x^{(0)} - y\|(1 + \frac{\tau}{2}F''(0))^k.$$

Therefore, we have exponential convergence when τ and $\|x^{(k)} - y\|$ are sufficiently small. □

Appendix 5. Additional numerical results

Comparison of filters M_n Figure 9 shows the Wasserstein error between the flow and the target for the SND kernel smoothed with $u = M_2$ and $u = M_4$. Here, we denote by **SND4** the smoothed negative distance $F := -\mathcal{I}_3[\text{abs} * u_\varepsilon]$ for $u_\varepsilon(x) = \frac{1}{\varepsilon}M_4(\frac{x}{\varepsilon})$.

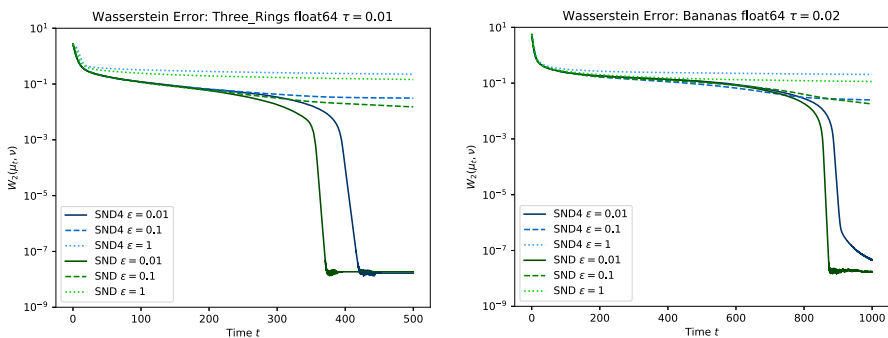


Fig. 9 W_2 error between target ν and flow γ_n^τ after n iterations. Horizontal axis in time $t = \tau n$. Both computed with double precision and step size $\tau = 0.01$ for the three-rings target (left) and $\tau = 0.02$ for the Bananas target (right)

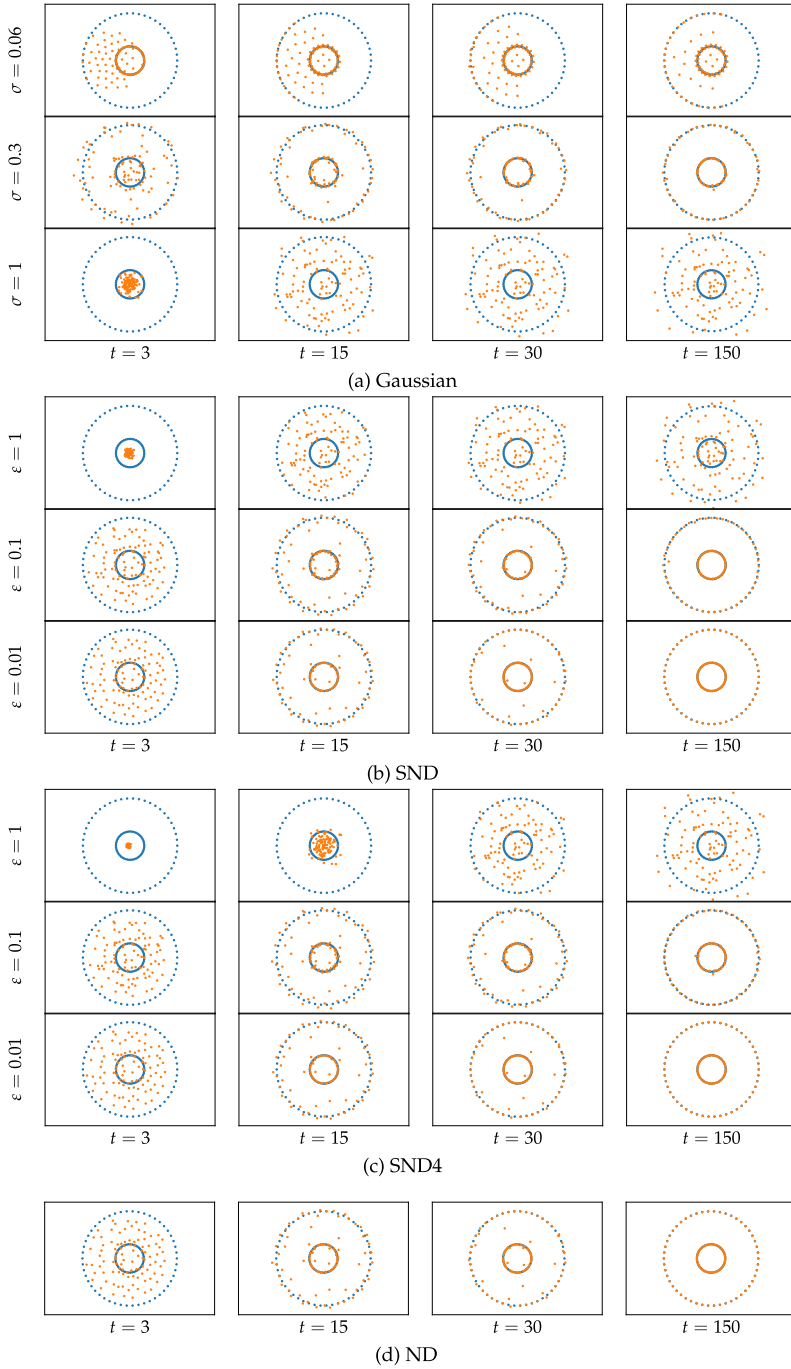


Fig. 10 MMD flow (26) with step size $\tau = 0.02$

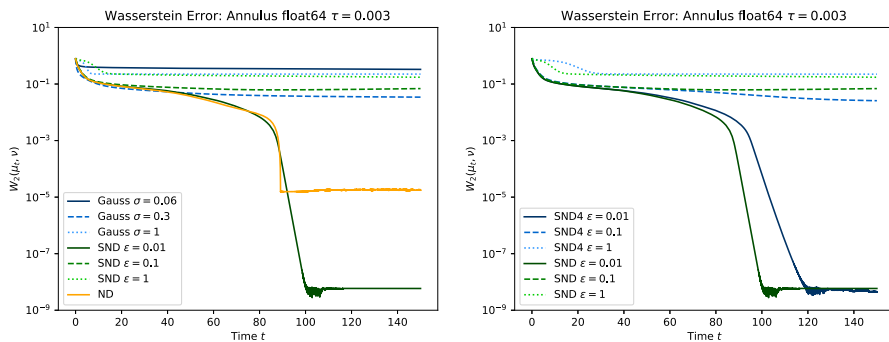


Fig. 11 W_2 error between Annulus target ν and flow γ_n^τ after n iterations. Horizontal axis in time $t = \tau n$. Both computed with double precision and step size $\tau = 0.003$

Table 1 Runtime in seconds for the Annulus target on a GPU for 50 000 gradient steps, averaged over 3 runs each

Kernel	Gauss	SND	ND	SND4
Time (s)	11.44	20.07	13.63	33.37

We keep the notation **SND** if we smooth with $u_\varepsilon(x) = \frac{1}{\varepsilon} M_2(\frac{x}{\varepsilon})$. Both SND and SND4 exhibit comparable error decay. Visually, the flows in Fig. 10b and c also show similar behavior. This suggests that the choice of the filter u has little impact on the behavior of the gradient flow. Note that using the same ε with M_4 or M_2 results in different smoothing strengths, as their supports differ. For large n , the derivation of $\mathcal{L}_3[\text{abs} * M_n]$ becomes increasingly tedious, and also, the numerical evaluation gets more expensive.

Annulus target The Annulus target consists of two concentric circles with radius 1 and 0.3. Each is discretized with 50 points, so that ν consists of $M = 100$ points. Here, we use a step size of $\tau = 0.003$ and double precision. The MMD flows are depicted in Fig. 10 and the respective errors in Fig. 11.

Computational times Table 1 provides an overview of the runtime of the four considered kernels for the Annulus target. The SND4 is significantly slower than the SND, due to the more complicated structure (see Example 4.8). However, as we saw above, it offers barely an advantage in accuracy.

Acknowledgements We thank Sebastian Neumayer for his valuable suggestions, especially in the numerical aspects of this work.

For open access purposes, the authors have applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

Author contribution All three authors Nicolaj Rux, Michael Quellmalz, and Gabriele Steidl have contributed equally to the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. MQs research was funded by the German Research Foundation (DFG): STE 571/19-1, project number 495365311, within the Austrian Science Fund (FWF) SFB 10.55776/F68 “Tomography Across the Scales.” GS acknowledges the funding

support by the DFG within the Excellence Cluster MATH+. NR gratefully acknowledges the funding support from the European Union and the Free State of Saxony (ESF).

Data availability Data availability is not applicable to this article as no new data were created or analyzed in this study.

Declarations

Informed consent Not applicable because no research involving humans has been conducted for this article.

Conflict of interest The authors declare no competing interests.

Research involving human and/or animals Not applicable because no research involving humans or animals has been conducted for this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abramowitz, M., Stegun, I.A.: Handbook of mathematical functions with formulas, graphs, and mathematical tables, 10th edn. Dover Publications, New York (1964)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows in metric spaces and in the space of probability measures, 2nd edn. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel (2008)
3. Arbel, M., Korba, A., Salim, A., Gretton, A.: Maximum mean discrepancy gradient flow. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32. Curran Associates (2019)
4. Balagué, D., Carrillo, J.A., Laurent, T., Raoul, G.: Dimensionality of local minimizers of the interaction energy. Arch. Ration. Mech. Anal. **209**, 1055–1088 (2013)
5. Bauschke, H.H., Combettes, P.L.: Correction to: Convex analysis and monotone operator theory in Hilbert spaces. Springer International Publishing, Cham (2017)
6. Borgwardt, K.M., Gretton, A., Rasch, M.J., Kriegel, H.-P., Schölkopf, B., Smola, A.J.: Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics* **22**(14), e49–e57 (2006)
7. Boufadène, S., Vialard, F.-X.: On the global convergence of Wasserstein gradient flow of the Coulomb discrepancy. HAL preprint hal-04282762 (2023)
8. Carrillo, J., Delgadino, M., Mellet, A.: Regularity of local minimizers of the interaction energy via obstacle problems. *Commun. Math. Phys.* **343**(3), 747–781 (2016)
9. Carrillo, J., Di Francesco, M., Esposito, A., Fagioli, S., Schmidtchen, M.: Measure solutions to a system of continuity equations driven by Newtonian nonlocal interactions. *Discrete Contin. Dynam. Systems* **40**(2), 1191–1231 (2020)
10. Carrillo, J., Huang, Y.: Explicit equilibrium solutions for the aggregation equation with power-law potentials. *Kinetic and Related Models* **10**(1), 171–192 (2017)
11. Carrillo, J., Shu, R.: From radial symmetry to fractal behavior of aggregation equilibria for repulsive-attractive potentials. *Calc. Var. Partial. Differ. Equ.* **62**(1) (2023)
12. Chafai, D., Saff, E.B., Womersley, R.S.: On the solution of a Riesz equilibrium problem and integral identities for special functions. *J. Math. Anal. Appl.* **515**, 126367 (2022)

13. Crisculo, G.: A new algorithm for Cauchy principal value and Hadamard finite-part integrals. *J. Comput. Appl. Math.* **78**, 255–275 (1997)
14. Cucker, F., Zhou, D.: *Learning theory: an approximation theory viewpoint*. Cambridge University Press (2007)
15. Duong, R., Stein, V., Beinert, R., Hertrich, J., Steidl, G.: Wasserstein gradient flows of MMD functionals with distance kernel and Cauchy problems on quantile functions. *ESAIM: COCV*, **32**(10), 37 (2026). <https://doi.org/10.1051/cocv/2025097>
16. Ehler, M., Gräf, M., Neumayer, S., Steidl, G.: Curve based approximation of measures on manifolds by discrepancy minimization. *Found. Comput. Math.* **21**(6), 1595–1642 (2021)
17. Flamary, R., Courty, N., Gramfort, A., Alaya, M.Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gautheron, L., Gayraud, N.T., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D.J., Tavenard, R., Tong, A., Vayer, T.: POT: Python optimal transport. *J. Mach. Learn. Res.* **22**(78), 1–8 (2021)
18. Frank, R.L., Matzke, R.M.: Minimizers for an aggregation model with attractive-repulsive interaction. *Arch. Ration. Mech. Anal.* **249**(15) (2025)
19. Gelfand, I., Shilov, G.: *Generalized functions*, vol. I. Academic Press, New York (1964)
20. Glaser, P., Arbel, M., Gretton, A.: KALE flow: a relaxed KL gradient flow for probabilities with disjoint support. In: *Advances in Neural Information Processing Systems*, vol. 34, pp. 8018–8031 (2021)
21. Gräf, M., Potts, D., Steidl, G.: Quadrature rules, discrepancies and their relations to halftoning on the torus and the sphere. *SIAM J. Sci. Comput.* **34**(5), 2760–2791 (2012)
22. Grafakos, L., Teschl, G.: On Fourier transforms of radial functions and distributions. *J. Fourier Anal. Appl.* **19**(1), 167–179 (2012)
23. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. *J. Mach. Learn. Res.* **13**(25), 723–773 (2012)
24. Gutleb, T.S., Carrillo, J.A., Olver, S.: Computation of power law equilibrium measures on balls of arbitrary dimension. *Constr. Approx.* **58**, 75–120 (2023)
25. Hagemann, P., Hertrich, J., Altekrieger, F., Beinert, R., Chemseddine, J., Steidl, G.: Posterior sampling based on gradient flows of the MMD with negative distance kernel. *Int. Conf. Learn. Representations (ICLR)* (2024)
26. Hertrich, J.: Fast kernel summation in high dimensions via slicing and Fourier transforms. *SIAM J. Math. Data Sci.* **6**, 1109–1137 (2024)
27. Hertrich, J., Gräf, M., Beinert, R., Steidl, G.: Wasserstein steepest descent flows of discrepancies with Riesz kernels. *J. Math. Anal. Appl.* **531**(1, Part 1), 127829 (2024)
28. Hertrich, J., Jahn, T., Quellmalz, M.: Fast summation of radial kernels via QMC slicing. *Int. Conf. Learn. Representations (ICLR)* (2025)
29. Hertrich, J., Neumayer, S.: Generative feature training of thin 2-layer networks. *Trans. Mach. Learn.* (2025)
30. Hertrich, J., Wald, C., Altekrieger, F., Hagemann, P.: Generative sliced MMD flows with Riesz kernels. In: *The Twelfth International Conference on Learning Representations (ICLR)* (2024)
31. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29**(1), 1–17 (1998)
32. Korba, A., Aubin-Frankowski, P., Majewski, S., Ablin, P.: Kernel Stein discrepancy descent. In: *Proceedings of the 38-th International Conference on Machine Learning (ICML)*, pp. 5719–5730. PMLR (2021)
33. Kraemer, F., Veselovska, A.: Enhanced digital halftoning via weighted sigma-delta modulation. *SIAM J. Imag. Sci.* **16**(3), 1727–1761 (2023)
34. Madych, W.R., Nelson, S.A.: Multivariate interpolation and conditionally positive definite functions. II. *Math. Comput.* **54**(189), 211–230 (1990)
35. Medhurst, R.G., Roberts, J.H.: Evaluation of the integral $I_n(b) = \frac{2}{\pi} \int_0^\infty (\frac{\sin x}{x})^n \cos(bx) dx$. *Math. Comput.* **19**(90), 123–126 (1965)
36. Micchelli, C.A.: Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constr. Approx.* **2**, 11–22 (1986)
37. Modeste, T., Dombry, C.: Characterization of translation invariant MMD on R^d and connections with Wasserstein distances. *J. Mach. Learn. Res.* **25**(237), 1–39 (2024)
38. Moreau, J.: Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France* **93**, 273–299 (1965)

39. Nemmour, Y., Kremer, H., Schölkopf, B., Zhu, J.-J.: Maximum mean discrepancy distributionally robust nonlinear chance-constrained optimization with finite-sample guarantee. In: 61st IEEE Conference on Decision and Control (CDC) (2022)
40. Nesterov, Y.: Lectures on Convex Optimization. Springer, Cham (2018)
41. Neumayer, S., Stein, V., Steidl, G., Rux, N.: Wasserstein gradient flows for Moreau envelopes of f -divergences in reproducing kernel Hilbert spaces. *Anal. Appl.* (2025)
42. Nüsken, N., Renger, D.M.: Stein variational gradient descent: many-particle and long-time asymptotics. *Found. Data Sci.* **5**(3) (2023)
43. Plonka, G., Potts, D., Steidl, G., Tasche, M.: Numerical Fourier Analysis, 2nd edn. Springer (2023)
44. Ragozin, D.L.: Rotation invariant measure algebras on Euclidean space. *Indiana Univ. Math. J.* **23**(12), 1139–54 (1974)
45. Rahimi, A., Recht, B.: Random features for large-scale kernel machines. *Adv. Neural Inf. Process. Syst.* **20** (2007)
46. Rux, N., Quellmalz, M., Steidl, G.: Slicing of radial functions: a dimension walk in the Fourier space. *Sampling Theory, Signal Processing, and Data Analysis* **23**(6) (2025)
47. Santambrogio, F.: Optimal Transport for Applied Mathematicians. Birkhäuser, Basel (2015)
48. Schmaltz, C., Gwosdek, P., Bruhn, A., Weickert, J.: Electrostatic half-toning. *Comput. Graph. Forum* **29**(8), 2313–2327 (2010)
49. Schoenberg, I.: Contributions to the problem of approximation of equidistant data by analytic functions. *Q. Appl. Math.* **4**, 45–99 (1946)
50. Sejdinovic, D., Sriperumbudur, B., Gretton, A., Fukumizu, K.: Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Stat.* **41**(5), 2263–2291 (2013)
51. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis, 4th edn. Cambridge University Press (2009)
52. Steinwart, I., Christmann, A.: Support Vector Machines. Information Science and Statistics. Springer, New York (2008)
53. Steinwart, I., Fasciati-Ziegel, J.: Strictly proper kernel scores and characteristic kernels on compact spaces. *Appl. Comput. Harmon. Anal.* **51**, 510–542 (2021)
54. Sun, X.: Conditionally positive definite functions and their application to multivariate interpolations. *J. Approx. Theory* **74**(2), 159–180 (1993)
55. Székely, G.: E-statistics: The energy of statistical samples. Technical Report, Bowling Green University (2002)
56. Villani, C.: Optimal Transport: Old and New. Springer, Berlin (2009)
57. Wendland, H.: Scattered Data Approximation. Cambridge University Press (2004)
58. Zhu, J.-J., Mielke, A.: Kernel approximation of Fisher–Rao gradient flows (2024). [Arxiv:2410.20622](https://arxiv.org/abs/2410.20622)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.