Received: 30 April 2024 | Revised: 12 June 2025 | Accepted: 5 January 2026

DOI: 10.1111/sjos.70054

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Spatial depth for data in metric spaces

Joni Virta

Department of Mathematics and Statistics, University of Turku, Turku, Finland

Correspondence
Joni Virta, Department of Mathematics and Statistics, University of Turku, Turku, Finland.
Email: joni.virta@utu.fi

Funding information
Academy of Finland, Grant/Award Numbers: 347501, 353769

Abstract

We propose a novel measure of statistical depth, the metric spatial depth, for data residing in an arbitrary metric space. The measure assigns high (low) values for points located near (far away from) the bulk of the data distribution, allowing quantifying their centrality/outlyingness. This depth measure is shown to have highly interpretable properties, making it appealing in object data analysis where standard descriptive statistics are difficult to compute. The proposed measure reduces to the classical spatial depth in a Euclidean space. In addition to studying its theoretical properties, to provide intuition on the concept, we explicitly compute metric spatial depths in several different metric spaces. Finally, we showcase the practical usefulness of the metric spatial depth in outlier detection, non-convex depth region estimation and classification.

KEYWORDS

distance function, object data, outlier detection, robustness, statistical depth

1 | INTRODUCTION

The purpose of this work is to propose a measure of *depth* for data residing in a general metric space $(\mathcal{X}, d)$. By depth we refer to a function that assigns to each point $\mu \in \mathcal{X}$ a measure of centrality $D(\mu; P)$ with respect to a given probability distribution P on $\mathcal{X}$. The larger $D(\mu; P)$ is, the more centrally located μ is with respect to the probability mass of P. And, vice versa,

points with small depths can be seen as outlying in the view of P. Our reason for working in a general metric space $\mathcal{X}$ is that many modern applications produce data that are inherently non-Euclidean (functions, compositions, trees, graphs, rotations, positive-definite matrices, etc.), known also as *object data*. Hence, devising methodology that works in arbitrary metric spaces allows for capturing all of these data types at once, see, for example, Bhattacharya and Patrangenaru (2003) and Lyons (2013) and Dubey and Müller (2019, 2022) and Dai and Lopez-Pintado (2023) and Virta et al. (2022) for a (highly non-exhaustive) list of works with a similar viewpoint. A particularly important aspect of object data analysis is interpretability (since the data spaces themselves are often quite unusual and exotic) and this plays a key role in the current work as well.

The study of depths in Euclidean spaces, that is, when $\mathcal{X} = \mathbb{R}^p$ and d is the Euclidean distance, has a long history. We do not attempt to paraphrase these developments here and refer the interested reader to Zuo and Serfling (2000) and Serfling (2006) and Mosler and Mozharovskyi (2022) instead. As a concrete example of a depth function, consider the *lens depth* (Liu & Modarres, 2011) $D_L(\mu; P)$ of a point μ in a Euclidean space with respect to a distribution P, defined as

$$D_L(\mu; P) := P(\| X_1 - X_2 \| \geq \max\{\| X_1 - \mu \|, \| X_2 - \mu \|\}), \tag{1}$$

where X_1, X_2 are drawn independently from P. That is, $D_L(\mu; P)$ gives the probability that the edge between X_1 and X_2 is the longest in the triangle formed by the points X_1, X_2, μ. This event can be understood as the point μ being "between" X_1 and X_2 and, hence, the greater its probability, the more central the point μ has to be. Other examples of depth functions include the halfspace depth (Tukey, 1975), Oja depth (Oja, 1983), the simplicial depth (Liu, 1990), β-skeleton depths (Yang & Modarres, 2018) and the spatial depth (Chaudhuri, 1996; Vardi & Zhang, 2000), which is also known as the L_1-depth and is treated in more detail below.

Inspection of the different depth measures reveals that they are typically heavily motivated by geometric arguments. As such, it is reasonable to expect that their underlying ideas would apply also if the data resides in an arbitrary metric space instead of a Euclidean space. This viewpoint for the lens depth was taken in Kleindessner and Von Luxburg (2017) and Cholaquidis et al. (2023) and Geenens et al. (2023), where it was shown that the definition (1) leads to a meaningful measure of depth for data residing in a general metric space $(\mathcal{X}, d)$ when the Euclidean distances in (1) are replaced with d. In Dai and Lopez-Pintado (2023), an equivalent treatment was given to halfspace depth. As a natural continuation to these works, the contribution of the current paper is to give a similar treatment to the classical spatial depth,

$$D_S(\mu; P) := 1 - \|\mathrm{E}\{\mathrm{sgn}(X - \mu)\}\|^2, \tag{2}$$

where $X \sim P$, $\mathrm{sgn}(x) := \mathrm{I}(x \neq 0)x/\| x \|$ and $\| \cdot \|$ denotes the Euclidean norm. Note that, strictly speaking, definition (2) does not equal the classical definition of L_1-depth/spatial depth (Vardi & Zhang, 2000) which does not take the square of the norm. As such, (2) and the classical spatial depth are a monotone transformation apart. Our reason for using a definition that slightly differs from the standard (but still retains exactly the same interpretation) is that this choice will greatly simplify the resulting expression for the metric space extension of D_S.

To give an idea of the contents of this paper, we next summarize the definition and the most interesting properties of our proposed metric space extension of (2). Let $(\mathcal{X}, d)$ be a complete and separable metric space and let P be a probability distribution on $\mathcal{X}$. Consider, for the sake of this demonstration, a simplified scenario where the distribution P has no atoms, that is, $P(X = \mu) = 0$

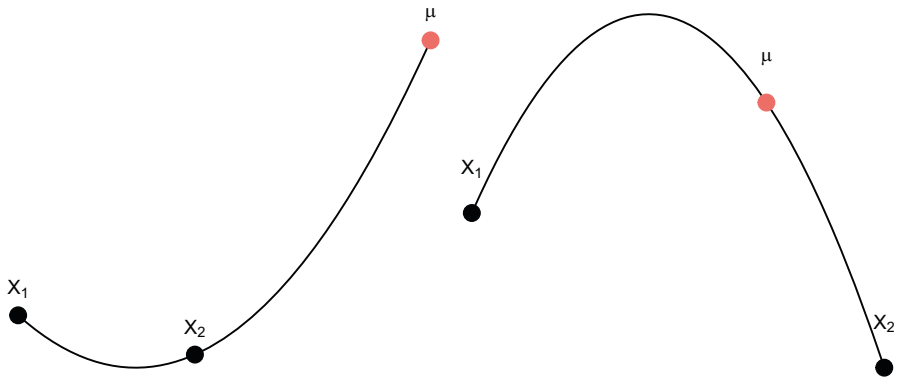

FIGURE 1 The left panel demonstrates the event $L[X_1, X_2, \mu]$ and the right panel the event $L[X_1, \mu, X_2]$ in an abstract case. The curved lines can be interpreted either as geodesics or "paths of shortest transformation" between two objects, depending on the viewpoint one takes.

for all $\mu \in \mathcal{X}$. In such a case, our proposed *metric spatial depth* of the point $\mu \in \mathcal{X}$ with respect to P takes the form

$$D(\mu; P) = 1 - \frac{1}{2}\mathrm{E}\left\{ \frac{d^2(X_1, \mu) + d^2(X_2, \mu) - d^2(X_1, X_2)}{d(X_1, \mu)d(X_2, \mu)} \right\}, \tag{3}$$

where X_1, X_2 are drawn independently from P. Although the connection between the two expressions is not obvious, a straightforward computation reveals that (3) reduces to (2) when $(\mathcal{X}, d)$ is a Euclidean space.

While, from definition (3) alone, it is not at all clear what $D(\mu; P)$ is measuring in the case of a general metric, it actually turns out that $D(\mu; P)$ captures the geometric structure of $\mathcal{X}$ in a highly natural way. To describe this behavior, we first define some notation. Given three points, $x_1, x_2, x_3 \in \mathcal{X}$, we denote by $L[x_1, x_2, x_3]$ the event that $d(x_1, x_3) = d(x_1, x_2) + d(x_2, x_3)$, that is, that equality is achieved in the triangle inequality with a specific permutation of the three points. One intuitive way to think about the event $L[x_1, x_2, x_3]$ is in terms of "object transformations." Namely, if we take $d(x_1, x_2)$ to denote the cost of "transforming" the object $x_1 \in \mathcal{X}$ into the object $x_2 \in \mathcal{X}$, then the triangle inequality says that a direct transformation is always at least as cheap as going through an intermediate object. Similarly, $L[x_1, x_2, x_3]$ is equivalent to saying that transforming x_1 to x_2 and then to x_3 is no more costly than directly transforming x_1 to x_3. For this to be possible, it is clear that x_2 has to, in some sense, reside in the middle of x_1 and x_3. In case the space $(\mathcal{X}, d)$ admits unique geodesics, then a simpler interpretation is available: $L[x_1, x_2, x_3]$ is equivalent to x_2 residing in the geodesic connecting x_1 and x_3, see the visualizations in Figure 1. Regardless of the space and interpretation, it is thus clear that the event $L[x_1, x_2, x_3]$ corresponds to x_2 being *centrally located* with respect to the other two points. In the sequel, we will use phrases such as "x_2 lies in between x_1 and x_3" to refer to the event $L[x_1, x_2, x_3]$.

Using the above terminology, our Theorem 3 in Section 2 says that the quantity $D(\mu; P)$ defined in (3) takes values in $[0, 2]$ and that:

$$D(\mu; P) = 0 \quad \text{if and only if} \quad P(L[X_1, X_2, \mu] \cup L[X_2, X_1, \mu]) = 1, \tag{4}$$

$$D(\mu; P) = 2 \quad \text{if and only if} \quad P(L[X_1, \mu, X_2]) = 1, \tag{5}$$

where X_1, X_2 are drawn independently from P. The events $L[X_1, X_2, \mu]$ and $L[X_1, \mu, X_2]$ have been visualized in Figure 1. These two extreme cases justify calling $D(\mu; P)$ a measure of depth: For a point μ to have maximal depth it must by (5) almost surely lie between any two points drawn from P. It goes without saying that, for this to happen, μ has to be a very central/deep point of P. Moreover, intuition says that such a point μ might not actually exist in every given metric space. Indeed, we later show that, when $(\mathcal{X}, d)$ is a Euclidean space, no point may have metric spatial depth equal to 2 and that, in fact, the maximal depth of a point in a Euclidean space is 1.

In a similar vein, for achieving the minimal depth, $D(\mu; P) = 0$, by (4) one of the events $L[X_1, X_2, \mu]$, $L[X_2, X_1, \mu]$ must always occur, in the almost sure sense. Intuitively, any point μ satisfying this must be located sufficiently outside of the main probability mass of P, making it an outlier. The condition is perhaps most easily understood in an extreme case where μ is exceedingly far away from the bulk of the distribution P. Then $d(X_1, X_2)$ is negligible compared to $d(X_1, \mu)$ and $d(X_2, \mu)$ and, by "zooming out," we see that in this case equality is approximately achieved in the triangle inequality, roughly satisfying the condition in (4) and implying that μ should have depth close to zero. This is indeed what happens, as is shown asymptotically later in Theorem 4 in Section 2. Finally, as pointed out earlier, properties such as (4) and (5) are highly desirable in object data analysis, where standard statistical intuition about Euclidean data no longer suffices.

Whereas the classical depth literature reviewed earlier targets exclusively Euclidean data, depth measures have also been developed for data living in specific non-Euclidean metric spaces. The most common of these is functional data (living in L_p-spaces), see the reviews in Nieto-Reyes and Battey (2016) and Gijbels and Nagy (2017). De Micheaux et al. (2021) propose depth for spatial curve data. Bachoc et al. (2024) develop spatial depth for density data in Wasserstein space. Their optimal-transport-based construction is tailored specifically for this space and, as a consequence, differs from our proposal (3) when $(\mathcal{X}, d)$ is the 2-Wasserstein space, see the comparison in (Bachoc et al., 2024, section 7.3). Liu and Singh (1992) and Pandolfo et al. (2018) define depths for data on the surface of a unit sphere (directional data). Chen et al. (2018) and Paindaveine and Van Bever (2018) extend the halfspace depth to scatter matrices (positive definite matrices), whereas Chau et al. (2019) propose several depths for positive definite matrix data using the Riemannian structure of the space. To summarize, each of these works focuses on one specific metric space (or family thereof), leading to effective ways of analyzing data in these particular spaces. Whereas, our approach is more general, and can be applied to any metric space. As with any object data method, this generality comes at the price of possibly less far-reaching conclusions, our depth being based on the metric space structure and unable to use any finer geometries of the spaces.

While our work is targeted toward data sets lying in non-Euclidean spaces (as, in an Euclidean space, the proposed concept reduces to the modified classical spatial depth (2)), it can still provide novel insights for Euclidean-valued data as well. Namely, as we demonstrate in Section 5, our proposed metric spatial depth allows combining a non-linear transformation with depth estimation for data in $\mathbb{R}^p$ in a natural way. To achieve this, we consider two alternatives, the kernel trick and graph distances. The first of these leads into a special case of our proposed method, where the metric space is a reproducing kernel Hilbert space, that was studied already in Chen et al. (2008). The non-linearization lets us, in particular, achieve depth regions that are non-convex and adapt to non-standard data shapes, see Figure 5 in Section 5. Similar ideas were used in the context of outlier detection in Schreurs et al. (2021). In the same vein, our final example in Section 5 shows how the metric spatial depth can be adapted to produce "L_p-versions" of the classical spatial depth (which is obtained when $p = 2$).

The main contributions of the current work are:

- We propose a completely novel extension of the spatial depth to arbitrary metric spaces and extensively study its robustness, interpretation, continuity and invariance properties. In particular, our results imply that the metric spatial depth is, in a sense, *more* natural concept than its Euclidean counterpart, in the sense that one must go to a general metric space to observe the full range of its behavior.
- To make the concept more accessible, we explicitly compute the metric spatial depth in four different scenarios, shedding further light on its intuitive meaning. The closed-form computations are made possible by the non-discrete nature of (3).
- We discuss the properties, including root-n-consistency, time complexity and possible computational shortcuts, of the sample metric spatial depth.
- We apply the metric spatial depth to three different practical scenarios: outlier detection, non-convex depth region estimation and classification, also comparing it to metric lens depth and metric halfspace depth in both accuracy and timing.

The manuscript is organized as follows. In Section 2, we formally define the metric spatial depth and study its theoretical properties in the case of an arbitrary metric space. Section 3 explores the metric spatial depth more closely in four particular metric spaces, whereas in Section 4 we discuss its sample estimation. Data examples are collected to Section 5 and in Section 6, we finally conclude with some discussion. The proofs of all technical results are collected to Appendix.

2 | METRIC SPATIAL DEPTH

Let $(\mathcal{X}, d)$ be a complete and separable metric space and let P be a probability distribution on $\mathcal{X}$. Throughout this work, we make the implicit assumption that every space is equipped with its Borel σ-algebra, guaranteeing, in particular, that all continuous functions are measurable. To define our proposed concept of depth, we first introduce the auxiliary function $h : \mathcal{X}^3 \to \mathbb{R}$,

$$h(x_1, x_2, x_3) := \mathbb{I}(x_3 \notin \{x_1, x_2\}) \frac{d^2(x_1, x_3) + d^2(x_2, x_3) - d^2(x_1, x_2)}{d(x_1, x_3) d(x_2, x_3)}, \tag{6}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Using h, we define the metric spatial depth $D(\mu; P)$ of the point $\mu \in \mathcal{X}$ with respect to P to be

$$D(\mu; P) := 1 - \frac{1}{2}\mathrm{E}\{h(X_1, X_2, \mu)\}, \tag{7}$$

where $X_1, X_2 \sim P$ are independent. The remainder of this section is devoted to studying the properties of $D(\mu; P)$ and we divide the treatment to four categories: robustness, interpretation, continuity and invariance.

2.1 | Robustness of $D(\mu; P)$

As with Euclidean statistical methods, also distance-based methodology commonly makes moment assumptions. In the latter case these are typically expressed as requirements of the form $\mathrm{E}\{d^k(X_1, X_2)\} < \infty$ where $X_1, X_2 \sim P$ are independent and $k \in \mathbb{N}$. For example, this condition

with $k = 2$ is in some sense equivalent to the requirement of finite variance (to which it reduces when d is the Euclidean distance in $\mathbb{R}$). The next result shows that the metric spatial depth $D(\mu; P)$ makes no such assumptions, making it a *robust* measure of depth, applicable to any distribution P. We use $\mathcal{P}$ to denote the set of all probability distributions taking values in $\mathcal{X}$.

Theorem 1. *$D(\mu; P)$ is finite for all $\mu \in \mathcal{X}$ and $P \in \mathcal{P}$.*

Another classical requirement for calling a statistic robust is that its *influence function* should be bounded (Hampel et al., 1986). The influence function $IF : \mathcal{X} \to \mathbb{R}$ of $D(\mu; P)$ at $z \in \mathcal{X}$ is defined as,

$$IF(z; D, \mu, P) := \lim_{\varepsilon \downarrow 0} \frac{D(\mu; (1-\varepsilon)P + \varepsilon\delta_z) - D(\mu; P)}{\varepsilon},$$

where $(1-\varepsilon)P + \varepsilon\delta_z$ is a mixture of P and δ_z, a Dirac point mass at $z \in \mathcal{X}$. Thus, IF measures the infinitesimal change in the metric spatial depth when the distribution P is contaminated by a small probability mass at point z. For D to be considered robust, the influence function must be bounded in z, that is, there must be an upper limit for the relative effect that any contamination, no matter how outlying, is able to cause to the depth of the point μ. Our next result reveals this to be the case.

Theorem 2. *For fixed $z \in \mathcal{X}$, we have,*

$$IF(z; D, \mu, P) = 2 - 2D(\mu; P) - \mathrm{E}\{h(X, z, \mu)\}.$$

Moreover,

$$\sup_{z \in \mathcal{X}} | IF(z; D, \mu, P) | \leq 4.$$

2.2 | Interpretation of $D(\mu; P)$

Recall next from Section 1 that $L[x_1, x_2, x_3]$ denotes the event that $d(x_1, x_3) = d(x_1, x_2) + d(x_2, x_3)$, that is, that the three points $x_1, x_2, x_3 \in \mathcal{X}$ are such that x_2 lies in the middle of the other two. Note also that this notation is not "unique" in the sense that, for example, $L[x_1, x_2, x_3] = L[x_3, x_2, x_1]$. We now have the following characterization about the range of possible values for $D(\mu; P)$.

Theorem 3. *$D(\mu; P)$ takes values in the interval $[0, 2]$. Additionally, letting X, X_1, X_2 denote independent draws from P,*

i. *$D(\mu; P) = 0$ if and only if*

$$P(X = \mu) = 0 \quad \text{and} \quad P(L[X_1, X_2, \mu] \cup L[X_2, X_1, \mu]) = 1.$$

ii. *$D(\mu; P) = 2$ if and only if*

$$P(X = \mu) = 0 \quad \text{and} \quad P(L[X_1, \mu, X_2]) = 1.$$

Part (i) of Theorem 3 says that for a point μ to have depth equal to zero, a necessary condition is that the distribution P must be such that $L[X_1, X_2, \mu] \cup L[X_2, X_1, \mu]$ always occurs. The standard case where this happens is when $\mathcal{X} = \mathbb{R}$ (meaning that any three points always fall onto a line) and P is supported on an interval. Any point μ outside of this interval then has depth equal to zero. Part (ii), on the other hand, says that maximal depth $D(\mu; P) = 2$ is achieved by any non-atom point μ such that, given two independent realizations X_1, X_2, the point μ always lies between these two points. This condition is very strict and can be satisfied only in some specific metric spaces veering toward pathological, see Section 3.2 for an example. Interestingly, in an Euclidean space, the maximal achievable depth is $D(\mu; P) = 1$, see Section 3.1.

Theorem 3 provides an interesting connection between the metric spatial depth and the metric lens depth (Cholaquidis et al., 2023; Geenens et al., 2023), defined as,

$$D_L(\mu; P) = P\big[d(X_1, X_2) \geq \max\{d(X_1\mu), d(X_2\mu)\}\big], \tag{8}$$

where $X_1, X_2 \sim P$ are independent. It is obvious that D_L takes values in the interval $[0, 1]$ and our next result connects these end points to the extremal behavior of the metric spatial depth. Its proof is a direct consequence of Theorem 3 and we omit it.

Corollary 1. *Let $\mu \in \mathcal{X}$ be fixed.*

i. *If $D(\mu; P) = 0$ then $D_L(\mu; P) = 0$.*
ii. *If $D(\mu; P) = 2$ then $D_L(\mu; P) = 1$.*

Corollary 1 essentially says that the range of behaviors that D considers absolutely extremal is a subset of the behaviors that D_L sees as absolutely extremal. For example, if the former marks a point μ as fully outlying, then so does the latter (but not necessarily vice versa), making metric spatial depth more "fine-grained" than the metric lens depth. We compare these two depths further later in the simulations of Section 5.

We next turn our attention to an asymptotic version of Theorem 3(i). Let μ_n be a sequence of points in $\mathcal{X}$. We say that μ_n *diverges* if $d(\mu_n, \mu) \to \infty$ as $n \to \infty$ for any fixed point $\mu \in \mathcal{X}$ (triangle inequality shows that if this condition holds for one point in $\mathcal{X}$, then it holds for all points). That is, divergent sequences are such that they eventually get arbitrary far from any fixed point and, naturally, the existence of such a sequence implies that the space $(\mathcal{X}, d)$ is unbounded.

Theorem 4. *Let μ_n be a divergent sequence in $\mathcal{X}$. Then $D(\mu_n; P) \to 0$ as $n \to \infty$.*

The result of Theorem 4 is very unsurprising when one considers it as a "limiting case" of part (i) of Theorem 3: when the diverging sequence μ_n gets farther and farther away from the bulk of the distribution P, the *relative* distance between X_1, X_2 diminishes (when compared to the growing $d(\mu_n, X_1)$ and $d(\mu_n, X_2)$), yielding an approximate equality in the triangle inequality and making the depth $D(\mu_n; P)$ thus approach the minimal value given in Theorem 3.

2.3 | Continuity properties of $D(\mu; P)$

As our next two results, we show that the depth D is continuous both with respect to the point μ and the distribution P. As the definition of D involves division by $d(X_1, \mu)d(X_2, \mu)$, these results require that P has no mass at the point $\mu \in \mathcal{X}$ in question. In the following, the notation $\rightsquigarrow$ denotes the weak convergence of probability distributions.

Theorem 5. *Fix $\mu \in \mathcal{X}$ and assume that $P(X = \mu) = 0$ for $X \sim P$. Let P_n be a sequence of probability distributions in $\mathcal{X}$ such that $P_n \rightsquigarrow P$ as $n \to \infty$. Then, as $n \to \infty$, we have*

$$D(\mu; P_n) \to D(\mu; P).$$

We demonstrate the use of Theorem 5 in Section 3.4 to compute the depths of points in the unit circle under the arc length distance through a limiting argument.

Theorem 6. *Fix $\mu \in \mathcal{X}$ and assume that $P(X = \mu) = 0$ for $X \sim P$. Let μ_n be a sequence in $\mathcal{X}$ such that $\mu_n \to \mu$ in the metric d as $n \to \infty$. Then, as $n \to \infty$, we have*

$$D(\mu_n; P) \to D(\mu; P).$$

Theorem 6 implies, in particular, that the metric spatial depth is continuous in the whole of $\mathcal{X}$ as long as P has no atoms. We omit the proof of Theorem 6 as the result follows from a direct application of Lebesgue's dominated convergence theorem. Later, when discussing the sample version of D in Section 4, we present still one more continuity result for the metric spatial depth (root-n consistency).

2.4 | Invariance properties of $D(\mu; P)$

We conclude the section by discussing invariance properties of the metric spatial depth. Assume that the metric d is invariant to a group $\mathcal{G}$ of transformations $g : \mathcal{X} \to \mathcal{X}$. That is, $d(gx_1, gx_2) = d(x_1, x_2)$ for all $g \in \mathcal{G}$, $x_1, x_2 \in \mathcal{X}$. Then, we clearly have $D(g\mu; P_{gX}) = D(\mu; P_X)$ for any $g \in \mathcal{G}$ where P_X and P_{gX} denote the distributions of X and gX, respectively. We give a few concrete examples: (i) If $(\mathcal{X}, d)$ is either the p-dimensional Euclidean space or the p-dimensional unit sphere equipped with the arc length metric, then the depth D is invariant (in the previous sense) to orthogonal transformations. (ii) If $\mathcal{X} = \{0, 1\}^p$ and d is the Hamming distance, then D is invariant to any bit-flip operations. (iii) If $\mathcal{X}$ is the space $\mathcal{S}^p$ of $p \times p$ positive-definite symmetric matrices and d is the corresponding Riemannian distance (Bhatia, 2009), then D is invariant to any transformations of the form $S \mapsto ASA'$ where $S \in \mathcal{S}^p$ and $A \in \mathbb{R}^{p \times p}$ is invertible.

Assume now that the random object X is such that $P_X = P_{gX}$ for all $g \in \mathcal{G}$. Then, by the previous paragraph, we have $D(g_1\mu, P_X) = D(g_2\mu, P_X)$ for all $g_1, g_2 \in \mathcal{G}$. That is, the metric spatial depth is constant on the orbit $\{g\mu | g \in \mathcal{G}\}$ for any point $\mu \in \mathcal{X}$. As an example of this, we continue our example scenario (i) above: any p-variate random vector X in a Euclidean space satisfying $X \sim OX$ for all $p \times p$ orthogonal matrices O is said to be spherical (Fang et al., 1990). As the orbits of the group of orthogonal transformations are origin-centered hyperspheres, the contours of the metric spatial depth D for a spherical distribution are thus sphere-shaped around the origin. For further discussion on invariance in conjunction with computational complexity, see Section 6.

3 | FOUR EXAMPLE SCENARIOS

In this section, we study the properties of the metric spatial depth $D(\mu; P)$ in four specific metric spaces. The purpose of these example scenarios is to (i) give further

insight about the behavior and interpretation of the metric spatial depth, and (ii) demonstrate that our proposed concept allows for closed-form solutions in several interesting cases.

3.1 | Hilbert space

Let $(\mathcal{X}, \langle\cdot\cdot\rangle)$ be a Hilbert space and let $\|\cdot\|$ and d be the norm and the metric induced by the inner product, respectively. The sign function $\mathrm{sgn} : \mathcal{X} \to \mathcal{X}$ is defined as $\mathrm{sgn}(x) = \mathrm{I}(x \neq 0)x/\|x\|$. Let X be a random element in $\mathcal{X}$ satisfying $\mathrm{E}\|X\| < \infty$. Then we define the expected value of X in the usual way through the Riesz representation theorem (Conway, 1990), as the unique element $\mathrm{E}(X) \in \mathcal{X}$ satisfying $\mathrm{E}\langle X, \mu\rangle = \langle \mathrm{E}(X), \mu\rangle$ for all $\mu \in \mathcal{X}$.

Lemma 1. *For a Hilbert space* $(\mathcal{X}, \langle\cdot\cdot\rangle)$, *the metric spatial depth takes the form*,

$$D(\mu; P) = 1 - \|\mathrm{E}\{\mathrm{sgn}(X - \mu)\}\|^2.$$

Lemma 1 carries with it several insights: (i) The expression $D(\mu; P)$ is based on the same key quantity as the classical L_1-depth (Vardi & Zhang, 2000). That is, the depth of a point μ is determined by the length of the expected value of a unit vector drawn from μ toward a point generated randomly from P, see Figure 2 for an illustration. The L_1-depth is also known as the spatial depth, see, for example, Serfling (2006), justifying naming our proposed concept the "metric spatial depth." (ii) For a distribution P on the real line (1-dimensional Euclidean space), the metric spatial depth is seen to further reduce to

$$D(\mu; P) = 1 - \{P(X > \mu) - P(X < \mu)\}^2,$$

that is, the depth of a point μ is in this case fully determined by its quantile level. As discussed in Geenens et al. (2023), this is desirable since distributions in $\mathbb{R}$ induce a natural concept of order (through their quantiles) and, hence, any reasonable measure of depth therein should be based on this order structure. (iii) In a Hilbert space, such as $\mathbb{R}^p$ equipped with the Euclidean inner product, the maximal achievable depth is 1. In other words, Hilbert spaces are "too structured" to allow reaching the upper half of the range of $D(\mu; P)$. By Lemma 1, the depth value 1 in Hilbert spaces is obtained if and only if μ is the center of the distribution P in the sense that the average of unit length vectors drawn from μ toward random X has length 0.

3.2 | British rail metric

As our next example, we consider a more pathological case where the theoretical maximum depth value 2 can be reached. Let $\mathcal{X} = \mathbb{R}^2$ be the plane and take the metric d to be the British rail metric,

$$d(x_1, x_2) = \mathbb{I}(x_1 \neq x_2)(\|x_1\| + \|x_2\|),$$

where $\|\cdot\|$ denotes the Euclidean norm. This metric space models the situation where all train trips in the Southern England have to go through London (located in the origin). Let next P be

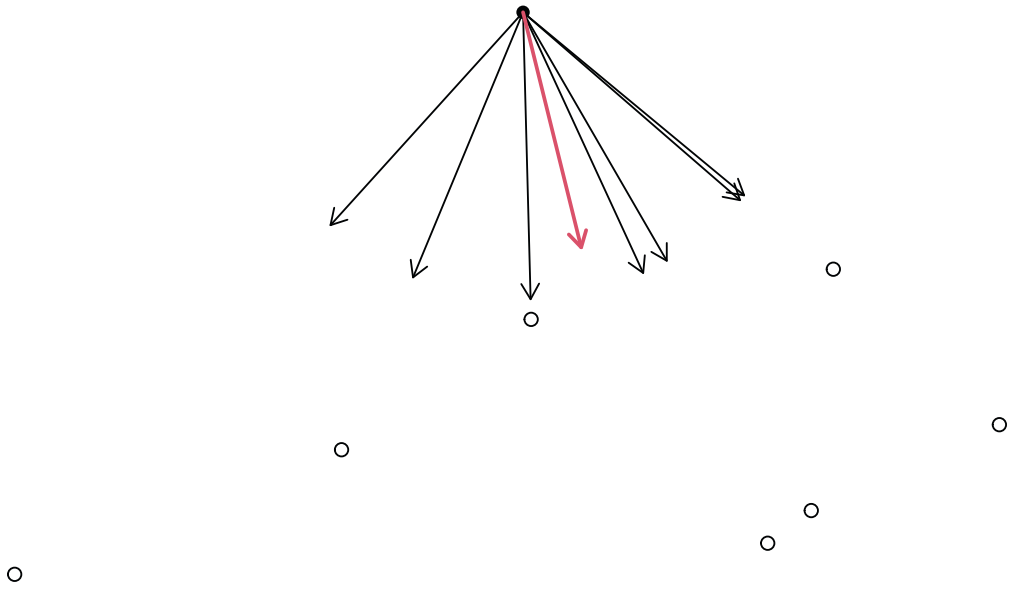

FIGURE 2 Assume that P puts equal mass to all seven circles shown in the plot. The black arrows are unit length vectors drawn from μ (the solid black point) toward these point masses. The red shorter arrow depicts the average of these vectors and its length (≈ 0.843) determines the depth of μ, larger values indicating smaller depths. As μ lies outside of the sample, a rather small depth, $D(\mu; P) \approx 0.289$, is now obtained.

any distribution in $\mathbb{R}^2$ having no atoms. Then $d(X_1, X_2)$ equals almost surely $\| X_1 \| + \| X_2 \|$ and the depth of a point $\mu \in \mathcal{X}$ takes the form

$$D(\mu; P) = \mathrm{E}\left\{ \frac{2 \| X_1 \| \| X_2 \|}{(\| X_1 \| + \| \mu \|)(\| X_2 \| + \| \mu \|)} \right\}, \tag{9}$$

showing that $D(0; P) = 2$.

The British rail metric appears also as the solution to the following optimality problem: Assume that P puts equal mass $1/n$ to each of n distinct points $x_1, \ldots, x_n \in \mathcal{X}$. Then, under what metric is $D(x_1; P)$ maximized? In other words, which combination of distances between n equiprobable distinct points makes x_1 as "central" as possible? The following lemma reveals that the answer is given by the British rail metric.

Lemma 2. *Let P put mass $1/n$ to each of n distinct points $x_1, \ldots, x_n \in \mathcal{X}$. Then,*

$$D(x_1; P) \le 1 + \left(1 - \frac{1}{n}\right)\left(1 - \frac{3}{n}\right),$$

with equality if and only if $d(x_i, x_j) = d(x_i, x_1) + d(x_1, x_j)$ for all distinct i, j in $\{2, \ldots, n\}$.

The equality condition in Lemma 2 corresponds to a situation where the shortest route between any two points always goes through x_1 ("London"). The actual distances between the points do not play a role in the solution, as long as they are positive. One particular, symmetric solution achieving the upper bound has been visualized in Figure 3. As is expected, the upper bound in Lemma 2 approaches the value 2 as $n \to \infty$.

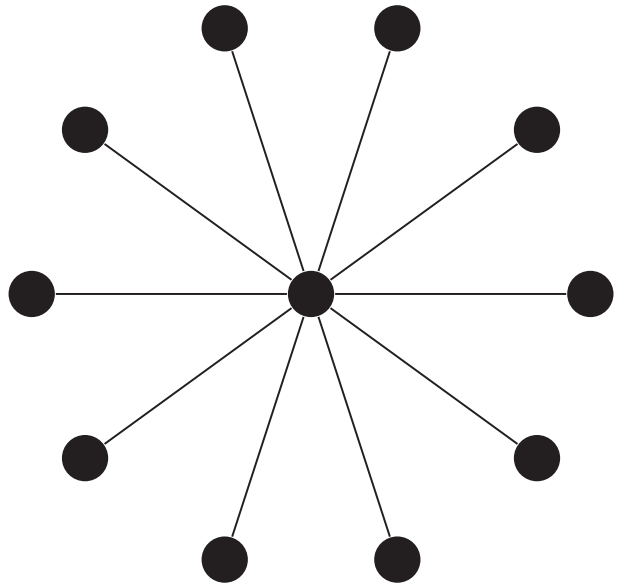

FIGURE 3 The picture displays a star graph on $n = 11$ points under which x_1 (the central point in the graph) achieves the maximal depth $D(x_1; P) = 2 - 41/121 \approx 1.661$. The distances in the graph are proportional to the edges, each of which has the same length.

3.3 | Discrete metric in a finite space

To gain further intuition on the metric spatial depth, we consider one of the simplest possible spaces, a finite set $\mathcal{X} := \{1,\dots,n\}$ equipped with the discrete metric $d(i,j) = \mathbb{I}(i \neq j)$. Any probability distribution P on $\mathcal{X}$ is then characterized by the probabilities $p_1,\dots,p_n$ of the n points. Our next result gives the values of the metric spatial depth in this scenario.

Lemma 3. *In a finite set* $\mathcal{X} = \{1,\dots,n\}$ *equipped with the discrete metric* $d(i,j) = \mathbb{I}(i \neq j)$, *we have*

$$D(1;P) = \frac{1}{2}\left(1 - \sum_{i=1}^{n} p_i^2\right) + p_1 \in [0,1].$$

Furthermore,

i. $D(1;P) = 0$ *if and only if exactly one of* $p_2,\dots,p_n$ *equals* 1.
ii. $D(1;P) = 1$ *if and only if* $p_1 = 1$.

Lemma 3 shows that the extremal depth values are obtained only in cases where the probability mass is concentrated solely at a single point. This is expected as, intuitively, the only way to make a point more outlying under a discrete metric where all distances are of equal length, is to reduce its probability mass. Similarly, as the concept of "location" carries no meaning under a discrete metric, the centrality of a point i is measured solely in the size of p_i relative to $p_1,\dots,p_{i-1},p_{i+1},\dots,p_n$.

Part (ii) of Lemma 3 says that if P gives a probability mass 1 to a point, then the depth of that point is exactly 1. A straightforward calculation shows that the same actually holds for the

metric spatial depth (7) in any metric space $(\mathcal{X}, d)$. Contrasting this observation with the results of Section 3.2 might lead one to ask that why should the scenario in Figure 3 lead to significantly higher depth values than are obtainable for points having probability mass 1. After all, it is difficult to imagine a "deeper" scenario than that of putting a Dirac point mass to a single point. The short answer is that there is no reason why the values $D(\mu; P)$ and $D(\mu; Q)$ should be directly comparable between two distributions P, Q (what *is* comparable are the values $D(\mu_1; P)$ and $D(\mu_2; P)$ for two different points μ_1, μ_2 under the same distribution P). As a longer answer, later in Section 6 we propose a modification of the metric spatial depth that better allows such comparisons.

3.4 | Uniform distribution on the unit circle

Our final example involves the metric space $\left(\mathbb{S}^1, d\right)$, studied commonly in directional statistics (Mardia & Jupp, 2000), where $\mathbb{S}^1$ is the unit circle in $\mathbb{R}^2$ and d is the arc length distance. That is, if the points $x_1, x_2 \in \mathcal{X}$ are expressed as two-dimensional unit vectors, the distance between them is $d(x_1, x_2) = \arccos\left(x_1' x_2\right)$. Plugging this in to (6) reveals that the resulting expression for $D(\mu; P)$ does not simplify in any meaningful way, making its theoretical analysis cumbersome. However, its exact values can still be computed in some specific cases and our next result gives a closed-form solution to the metric spatial depth under the assumption of a uniform distribution on $\mathbb{S}^1$. Its proof relies on constructing a sequence of discrete uniform distributions on the circle and invoking Theorem 5.

Theorem 7. *Let P be the uniform distribution on the unit circle $\mathbb{S}^1$. Then, for any point $\mu \in \mathbb{S}^1$, we have,*

$$D(\mu; P) = \pi^2/6 - 1 \approx 0.6449.$$

4 | SAMPLE METRIC SPATIAL DEPTH

Let $X_1, \ldots, X_n$ be a random sample from the distribution P. Denote the empirical distribution of the sample by P_n. The sample metric spatial depth $D(\mu; P_n)$ of the point $\mu \in \mathcal{X}$ is given by

$$D(\mu; P_n) = 1 - \frac{1}{2n^2} \sum_{i,j=1}^{n} h\left(X_i, X_j, \mu\right), \tag{10}$$

where h is defined as in (6). By isolating the terms with $i \neq j$ in the double sum, we see that $D(\mu; P_n)$ is asymptotically equivalent to a second order U-statistic, guaranteeing that $D(\mu; P_n)$ is a root-n-consistent estimator of the population depth $D(\mu; P)$.

Theorem 8. *For a fixed $\mu \in \mathcal{X}$, we have,*

$$D(\mu; P_n) = D(\mu; P) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),$$

as $n \to \infty$.

The U-statistic theory (Lee, 1990) further guarantees that $\sqrt{n}\{D(\mu;P_n) - D(\mu;P)\}$ admits a limiting normal distribution. Pursuing such finer asymptotic properties of $D(\mu;P_n)$ would allow, for example, testing hypotheses of the form $H_0 : D(\mu;P) = \alpha$, for some point μ and depth level $\alpha \in [0,2]$. However, such tests are rather impractical in the general situation as the range of $D(\mu,P)$ depends on the metric space in question, see Section 3.1 for an example. Thus, to test, for instance, whether $\mu \in \mathcal{X}$ is the deepest point w.r.t. P necessarily requires knowing the maximal value of the map $\mu \mapsto D(\mu;P)$. Consequently, in this work, to stay as general as possible, we have chosen not to study such questions.

In practice, for a realized sample $X_1,\ldots,X_n$, computing $D(\mu;P_n)$ for a single value of μ requires two nested loops over the sample, leading to time complexity of order $\mathcal{O}(n^2)$. As there does not appear to exist a straightforward connection between the terms involved in computing, for example, $D(X_1,P_n)$ and $D(X_2,P_n)$, finding the depths of the full sample is thus an $\mathcal{O}(n^3)$-operation. Such complexities are rather standard for methods based on distance matrices, see, for example, Cholaquidis et al. (2023) and Dai and Lopez-Pintado (2023). When compared to standard Euclidean data, object data sets often have comparatively smaller sample sizes, consider, for example, the data examples in Dai and Lopez-Pintado (2023) and Geenens et al. (2023), meaning that the $\mathcal{O}(n^3)$ complexity is usually acceptable in practice. However, faster computation is still possible in specific spaces. For example, for data in a Hilbert space (Section 3.1), computing the depths of a full sample using Lemma 1 is an $\mathcal{O}(n^2)$-operation. Whereas, under a discrete metric (Section 3.3) the depth of a point is by Lemma 3 essentially determined by its point mass, making the computation of all depths possible in $\mathcal{O}(n)$ time. Similar shortcuts are not possible in our two remaining examples (Sections 3.2 and 3.4) where computation of the depth of a single point requires two nested loops, leading to the full complexity of $\mathcal{O}(n^3)$, see (9) and the discussion in Section 3.4. Further computational savings could be obtained, at the expense of accuracy, by considering in the double sum in (10) only a subset of all possible n^2 pairs, as when using incomplete U-statistics (Blom, 1976). However, in the current work where our sample sizes are still manageable enough, we have exclusively used the full formula (10).

5 | EXAMPLES

We next exemplify the use of metric spatial depth in three tasks: outlier detection, non-convex depth region estimation and classification. The used methods are implemented in `R` and their program codes are available on the author's webpage, https://users.utu.fi/jomivi/software.

5.1 | Outlier detection

In our outlier detection task, we study how well different methods are able to separate outliers from the bulk of a distribution. We generate 10-dimensional points as $x_i \sim \mathcal{N}_{10}\big(\lambda 1_{10}, \mathrm{diag}(I_{10})\big)$ and transform them as $x_i \mapsto x_i / \| x_i \|$, meaning that the data lies on the unit sphere in $\mathbb{R}^{10}$. To capture the geometry of the space, we use the arc length distance as our metric. We create two groups, the bulk of size $(1-\varepsilon)n$ and the outlier group of size εn where the value $-\lambda$ is used for the location of the outlying group in place of λ (i.e., the bulk and the outlier group reside on the opposite sides of the sphere). We consider three different values of $\lambda = 1/4, 1/3, 1/2$, two different sample sizes, $n = 50,100$ and a range of different proportions of outliers, $\varepsilon = 0.01,0.02,\ldots,0.30$. A similar setting was used in Heinonen et al. (2025).

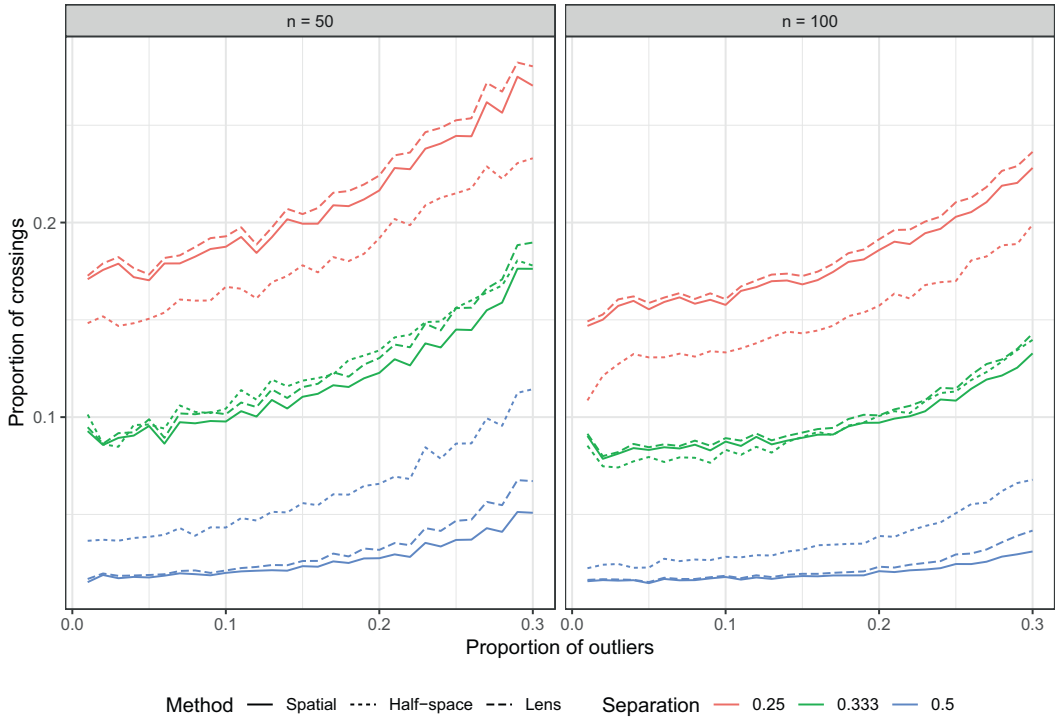

FIGURE 4 Proportions of bulk-outlier crossings C (see the main text for the definition) for the metric spatial, halfspace and lens depths under different settings over 1000 replications. A lower value indicates a better separation.

We conduct 1000 replicates of each setting and in each case compute our metric spatial depth (7), the metric lens depth (8) (Cholaquidis et al., 2020; Geenens et al., 2023) and the metric halfspace depth (Dai & Lopez-Pintado, 2023), defined as

$$D_H(\mu; P) := \inf_{z_1, z_2 \in \mathcal{X}, d(z_1,\mu) \leq d(z_2,\mu)} P\{d(X, z_1) \leq d(X, z_2)\}.$$

We compute D_H using Algorithm 1 in Dai and Lopez-Pintado (2023) using all sample pairs as anchor points. All three competitors output the set of depth values for the full sample and, for each set, we compute the average number C of "bulk-outlier-crossings," that is, the proportion of pairs of a bulk observation and an outlying observation such that the former gets a lower depth than the latter. Thus, the smaller C is, the better the depths identify/separate the outlying observations from the bulk. The average values of C for each combination of settings and method are shown in Figure 4.

From Figure 4 we make the following conclusions: (i) The performances of the spatial and lens depth are in all scenarios very close to each other. Based on Corollary 1 this is somewhat as expected. Moreover, the spatial depth is systematically the better one out of the two. We believe that this is connected to the fact that, by Theorem 3, low values of the metric spatial depth characterize a very natural form of outlyingness, something which is not guaranteed for the metric lens depth. (ii) As is natural, independent of the method, separation gets easier the larger the sample

TABLE 1 Average computation times of the three methods (in seconds) as a function of n.

Method	$n = 50$	$n = 100$	$n = 150$	$n = 200$
Spatial	0.022	0.147	0.476	1.106
Half-space	0.015	0.082	0.235	0.533
Lens	0.023	0.194	0.631	1.504

size is, the less we have outliers and the greater the separation between the groups is. (iii) The best two methods are the spatial depth and the halfspace depth. The spatial depth has an edge when the bulk and the outliers are sufficiently separated, whereas the halfspace depth dominates when the groups are close to each other. In both of the extreme cases $\lambda = 1/2, 1/4$, the difference between the two methods is rather large. Again we believe that this difference can be accounted for Theorem 3: large values of the metric spatial depth indicate a very specific form of centrality, which is likely ill-suited to this particular scenario, meaning that the method encounters difficulties when the outliers are brought closer to the center of the bulk.

The average computation times of the three methods over 100 replicates are shown in Table 1 as a function of n, when $\lambda = 1/4$ and $\epsilon = 0.1$ (note that the latter two parameters have no effect on the computation time). The computation of the distance matrix was not included in the times as this step is identical for all methods. Metric half-space depth is the fastest of the three to compute but overall there are no large differences between the methods. As claimed, all times are seen to scale approximately as $\mathcal{O}(n^3)$. All methods were implemented in base `R` and were timed with the `microbenchmark`-package (Mersmann, 2024) on a desktop computer with AMD Ryzen 53,600 6-core processor having 16 GB RAM.

5.2 | Depth and non-convex distributions

A common feature of several classical measures of depth in $\mathbb{R}^p$ is that they produce convex depth regions (Serfling, 2006). Naturally, this is undesirable as soon as the data exhibits non-convex shapes. In this section, we show how the proposed metric spatial depth can be adapted to obtain meaningful measures of depth also in such scenarios.

For illustration purposes, we use a simple bivariate dataset of size $n = 150$, generated as $x_i = 2(\cos(\theta_i), \sin(\theta_i)) + 0.1^{1/2}\epsilon$, where θ_i are i.i.d. from the uniform distribution on $[0, 2\pi]$ and ϵ are i.i.d. from the bivariate standard normal distribution. This yields a sample resembling a circle around the origin on the plane, see the black points in the panels of Figure 5. To exemplify the behavior of standard measures of depth under such a non-convex scenario, we divided the plane into a fine grid and computed the metric spatial depth of each grid point under the Euclidean metric. As described in Section 3.1, this yields depths equivalent to the classical L_1-depth. The contour plot of the depths, shown in the upper left panel of Figure 5 (yellow color indicates large values), reveals that the depths indeed fail to capture the natural shape of the sample, assigning the highest depth around the origin where no sample points lie. To obtain better performance, we next propose two alternative approaches.

As a first option, we combine metric spatial depth with the idea behind ISOMAP (Tenenbaum et al., 2000), a classical method of manifold learning. ISOMAP works by using only local Euclidean distance information and disregarding the longer distances, which usually fail to capture any curved structure in the data. Namely, for each point i in turn, we "erase" all but the k smallest distances $d(x_i, x_j), j \in \{1, \ldots, i-1, i+1, \ldots, n\}$, where k is a user-chosen parameter, and

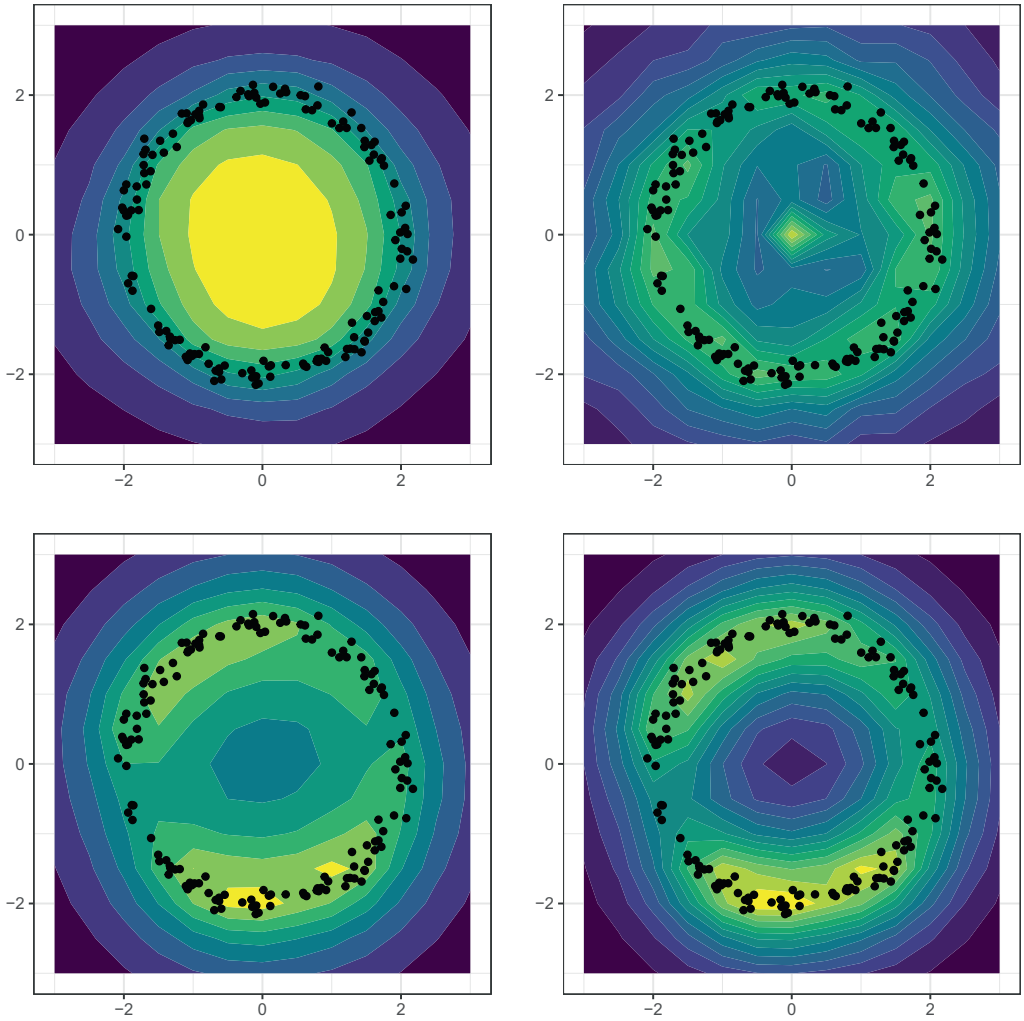

FIGURE 5 The depth contours of a circular data set of size $n = 150$ (the black points) under four different choices of distance for the metric spatial depth. The distances correspond, from left to right and from top to bottom: Euclidean distance, ISOMAP, RKHS with the rational quadratic kernel, RKHS with the Gaussian kernel. Yellow shades indicate the highest depths and the color scale is not uniform in the four panels.

use the remaining distances to construct a weighted undirected graph $\mathcal{G}$ on the n points. Next, we let $d^*(x_i, x_j)$ denote the graph distance between x_i, x_j in $\mathcal{G}$ (i.e., the length of the shortest path between x_i, x_j in the graph). The corresponding distances $d^*(x_i, \mu)$ between the sample and a "test point" μ are computed as graph distances after grafting μ to the graph $\mathcal{G}$ by its k nearest neighbors. For small k this results into d^* that approximates the geodesic distances (assuming the points lie roughly on some manifold, as in our example). However, k must also be kept large enough such that the graph $\mathcal{G}$ stays connected. The depth contours obtained with this procedure using $k = 8$ are shown in the upper right panel of Figure 5. We observe that the contours capture the circular shape of the data, but assign large depths also close to the origin. This happens because the origin is roughly equally far away from all sample points, and its k nearest neighbors are likely to

be roughly uniformly distributed around the circle. Consequently $d^*(x_i, 0)$ cannot be too large for any i, giving the origin a large depth.

As a second alternative, we resort to the well-known "kernel trick" and Moore–Aronszajn theorem (Aronszajn, 1950) which says that, for every symmetric, positive definite kernel $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, there exists a reproducing kernel Hilbert space (RKHS) $(\mathcal{H}, \langle \cdot\cdot \rangle)$, such that $\kappa(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle$ for some "feature mapping" $\phi : \mathcal{X} \to \mathcal{H}$ determined implicitly by the choice of κ. Now, every Hilbert space is a metric space, meaning that the corresponding metric can be expressed through the kernel as $d^2(x_1, x_2) = \kappa(x_1, x_1) - 2\kappa(x_1, x_2) + \kappa(x_2, x_2)$. Consequently, applying our proposed concept to these distances allows us to compute the metric spatial depths of the features $\phi(x_i)$. As the implicit mapping ϕ is typically non-linear, this allows capturing non-convexities in the original space $\mathcal{X}$. This special case of metric spatial depth was considered already in Chen et al. (2008).

As an example, we have used the rational quadratic kernel $\kappa(x_1, x_2) = \left(1 + \| x_1 - x_2 \|^2\right)^{-1}$ (the bottom left panel of Figure 5) and the Gaussian kernel $\kappa(x_1, x_2) = \exp\left(-0.933 \| x_1 - x_2 \|^2\right)$ (the bottom right panel), where $\| x_1 - x_2 \|$ denotes the Euclidean distance. Inspecting the results shows that both kernels manage to capture the circular shape of the data and produce satisfactory contours. While the rational quadratic kernel still puts some depth in the origin, the Gaussian kernel, whose tuning parameter value 0.933 has been chosen manually, obtains contours that assign minimal depth to the origin. If depths were used as a part of some supervised learning technique, standard cross-validation could also be used to select any tuning parameter values involved in the distance function, see the next section for an example.

5.3 | Classification

As our third example, we apply the metric spatial depth to a classification problem through the technique of depth–depth (DD) classification (Li et al., 2012). Let $x_1, \ldots, x_{n_1} \in \mathcal{X}$ and $x_{n_1+1}, \ldots, x_{n_1+n_2} \in \mathcal{X}$ denote samples from two groups and denote the corresponding empirical distributions by P_{n1} and P_{n2}. In DD-classification, we compute the depth vectors $z_i := (D(x_i P_{n1}), D(x_i P_{n2}))$, $i \in \{1, \ldots, n_1 + n_2\}$ and fit a classification rule (such as a linear discriminant) to the two-dimensional sample $z_1, \ldots, z_{n_1+n_2}$. A test point $x \in \mathcal{X}$ is then classified based on the vector $z := (D(xP_{n1}), D(xP_{n2}))$. This procedure extends to multiple groups in an obvious way. The main idea behind the DD-classifier is that depth functions automatically take into account the geometry of the data, avoiding the need to estimate any distributional parameters and making the method fully non-parametric.

In this illustration, we use the Fashion MNIST dataset (Xiao et al., 2017) available at https://www.kaggle.com/datasets/zalando-research/fashionmnist. The data consists of 28×28 grayscale images of various clothing items grouped into 10 classes. For simplicity, we consider only the test part of the full data set and only three classes (dress, shirt, ankle boot). This gives us a working data set of $n = 3000$ observations of dimensionality 784. The first 80 images in this set are shown in Figure 6. We conduct a total of 100 replicates of our study and in each replicate draw randomly 150 training images and 50 test images from among the 3000 observations. We use DD-classification with the metric spatial depth to classify the test images and use as our criterion the proportion of correct classifications in the test set. As a classifier we use either LDA or QDA. The main purpose of this illustration is to further demonstrate that while metric spatial depth is mainly targeted toward non-standard forms of data, it can still be used to enhance also the analysis of classical Euclidean data. We achieve this by choosing as our metric the L_p-distance

FIGURE 6 The first 80 images in the used data. The data consists of three classes: dresses, shirts and ankle boots.

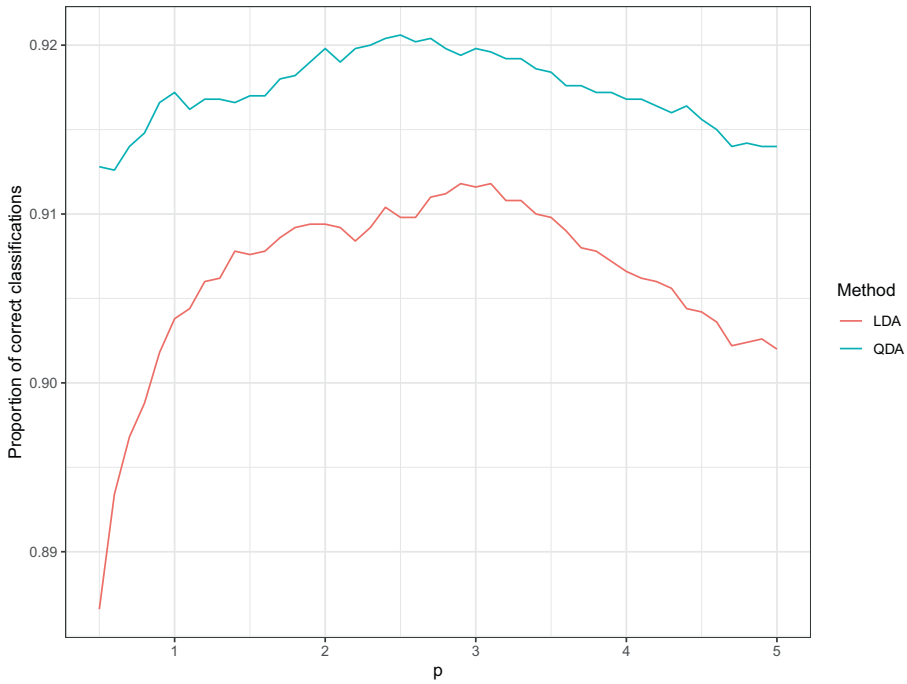

FIGURE 7 Average proportions of correct classifications as a function of the parameter p of L_p-distance. The two curves correspond to the different classification methods.

with different values of $p = 0.5,0.6,\ldots, 5$. The choice $p = 2$ thus corresponds to using the standard spatial depth (Chaudhuri, 1996; Vardi & Zhang, 2000) in the analysis and the remaining choices correspond to novel methodology. See also Perlibakas (2004) and Rodrigues (2018) for earlier uses of L_p-distances in the context of classification.

The lines in Figure 7 depict the average proportions of correct classifications over the replicates as a function of p. We make two observations of interest: (i) The mean curves appear to be rather smooth functions of p that achieve their maximum somewhere around $p \approx 3$ for LDA and $p \approx 2.5$ for QDA. In particular, for both LDA and QDA the maximum is achieved with a super-Euclidean geometry (L_p with $p > 2$). As the extreme case $p = \infty$ corresponds to using only a single pixel's worth of information, this probably means that the classification information in the original images is to an extent concentrated to a small group of pixels. (ii) While

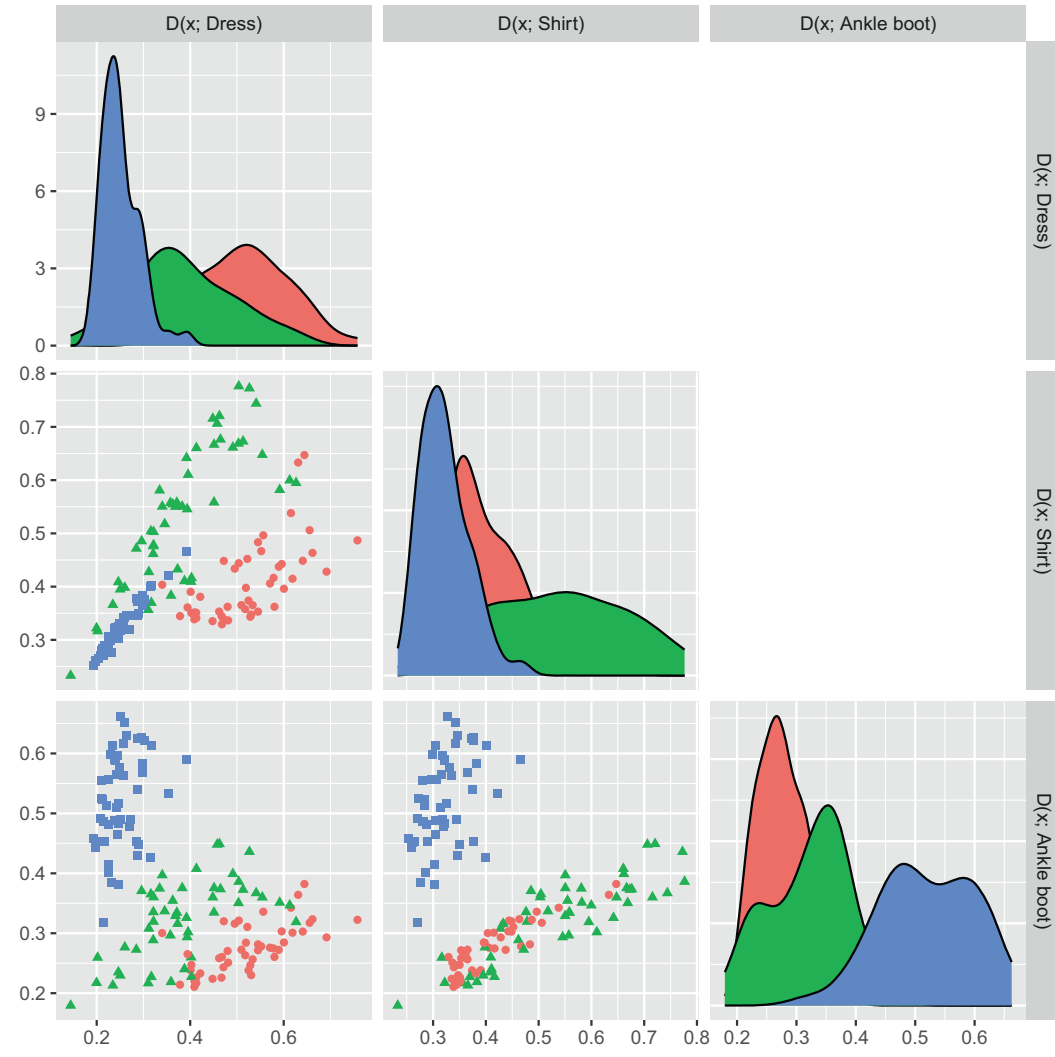

FIGURE 8 Paired scatter plots and histograms of the components of z_i for one replicate of the study with the L_3-norm. The colors and shapes correspond to the three classes: dress (red dot), shirt (green triangle), ankle boot (blue square).

the difference between LDA and QDA is minor, the latter appears the superior choice regardless of p, implying that the scatter plot of the points $z_i \in \mathbb{R}^3$ is not fully linearly separable. This is confirmed by Figure 8 where we plot the paired scatter plot of the 3-dimensional training points z_i for one replicate of the study with the choice $p = 3$. It is evident that while the separation between the groups is rather good, the point clouds still exhibit some curvature that make LDA suboptimal. Note also that the histograms on the diagonal of Figure 8 reveal that the largest values (that is, depths) in each component of z_i are indeed obtained by images of the corresponding class, as expected. To conclude, this (and the previous) example go to show that procedures targeted toward general object data can still benefit also the analysis of regular Euclidean data.

6 | DISCUSSION

We begin the discussion by commenting on the seeming discrepancy noted earlier in Section 3.3. For example, under the star-shaped graph shown in Figure 3, the central point x_1 has metric spatial depth $D(x_1; P) = 1$ when P charges its full probability mass to x_1, whereas its depth is much larger, $D(x_1; P) \approx 1.661$, when P is uniform on the nodes of the graph. As stated earlier, this happens simply because depth values under different distributions P might not be directly comparable. However, as such comparisons can sometimes be desired or even unavoidable, we next propose a modification of our main concept which avoids the previous "paradox." We define the modified metric spatial depth $D^*(\mu; P)$ as

$$D^*(\mu; P) \coloneqq D(\mu; P) + P(X = \mu)\{2 - P(X = \mu)\}.$$

That is, $D^*(\mu; P)$ is otherwise equal to $D(\mu; P)$, but points μ with positive probability mass have suitably increased depth. The form of the extra term $P(X = \mu)\{2 - P(X = \mu)\}$ is chosen precisely as it makes the modified metric spatial depth satisfy Theorem 9 below. Its proof is, up to conditioning, equivalent to the proof of Theorem 3 and, as such, omitted.

Theorem 9. *$D^*(\mu; P)$ takes values in the interval $[0, 2]$. Additionally, letting X, X_1, X_2 denote independent draws from P,*

i. *$D^*(\mu; P) = 0$ if and only if*

$$P(X = \mu) = 0 \quad \text{and} \quad P(L[X_1, X_2, \mu] \cup L[X_2, X_1, \mu]) = 1.$$

ii. *$D^*(\mu; P) = 2$ if and only if*

$$P(L[X_1, \mu, X_2] | X_1 \neq \mu, X_2 \neq \mu) = 1.$$

The interesting part of Theorem 9 is (ii) which says that the maximal depth value of 2 is reached if and only if, *conditionally on* $X_1 \neq \mu, X_2 \neq \mu$, the event $L[X_1, \mu, X_2]$ occurs almost surely. Hence, any probability mass at μ directly contributes to its depth and, for the maximal depth value 2 to occur, the restriction of P outside of $\{\mu\}$ should obey our earlier discussion surrounding formula (5). Note that in the pathological case where $P(X = \mu) = 1$, the conditional expectation in Theorem 9 is considered to equal 1, meaning that the maximal depth obtained by D^* under a point mass distribution is indeed 2. Examination of the proofs of Theorems 1, 2, 4, 5, 6 reveals that analogous results can be derived also for D^* (trivially in the case of the latter two). However, despite this, we still kept D as the main concept of this work for the following reasons: (i) the modification used to obtain D^* from D is somewhat artificial and not backed up by anything else besides performance at large point masses, (ii) the modified version loses the natural connection to the original spatial depth described in Section 3.1, and (iii) the change has little effect in most practical situations where no point anyway has mass greater than $1/n$.

We close the paper by commenting on possible avenues for further research. As with any measure of depth, a particularly interesting question regarding the metric spatial depth is finding the maximizers of the map $\mu \mapsto D(\mu; P)$ for a given probability distribution P. These maximizing points are typically called "medians" and, as location estimation plays an important role in almost all statistical practice, finding a way to carry out the maximization of $\mu \mapsto D(\mu; P)$ would

be practically very valuable. For example, the clustering method k-means is based solely on finding centers (location estimates) of would-be clusters and any concept of "robust metric location" can be used to formulate a robust metric version of k-means.

When $(\mathcal{X}, d)$ is an Euclidean space, the points with maximal metric spatial depth are precisely those $\mu \in \mathcal{X}$ which satisfy $\mathrm{E}\{\mathrm{sgn}(X-\mu)\} = 0$, see Section 3.1. The left-hand side of this condition is, under suitable regularity conditions for X, the gradient of the function $\mu \mapsto \mathrm{E} \parallel X - \mu \parallel$, showing that in the Euclidean case the spatial median is actually the L_1-equivalent of the mean vector. Based on this connection, it is natural to ask whether the same holds also in some other metric space. That is, are the minimizers of $\mu \mapsto \mathrm{E}\{d(X,\mu)\}$ such that they also maximize $D(\mu; P)$ for some non-Euclidean $(\mathcal{X}, d)$?

Finding the deepest points in practice, that is, maximizing $\mu \mapsto D(\mu; P_n)$ is in the Euclidean case typically handled through gradient-based optimization, see Vardi and Zhang (2000). However, such techniques do not work in a general metric space due to the lack of vector space structure. In some special cases, alternative structural properties of $\mathcal{X}$ can possibly be used instead (e.g., geodesic convexity in case $\mathcal{X}$ admits a representation as a manifold). If the maximizer is searched only among the sample $x_1, \ldots, x_n$ (instead of over the full space $\mathcal{X}$), then, as described in Section 4, determining the deepest of these points is a simple enumeration task of order $\mathcal{O}(n^3)$. Given a sample of sufficient size, the deepest sample point can likely be used as a reasonable approximation to the actual deepest point.

Compared to other Euclidean depths, the classical spatial depth can be thought to offer a compromise between invariance properties and computational complexity: it is both robust and fast to compute in practice but lacks affine invariance which is often desired from depths (Zuo & Serfling, 2000). As such, a natural question is whether the metric spatial depth is subject to a similar trade-off. The answer to this depends highly on the space in question. Take, for example, the space $(\mathcal{X}, d)$ of $p \times p$ positive definite matrices equipped with the Riemannian metric $d(X_1, X_2) = \parallel Log\left(X_1^{-1/2} X_2 X_1^{-1/2}\right) \parallel_F$. The most natural class of transformations in $\mathcal{X}$ are the similarity transformations $X \mapsto AXA'$, where $A \in \mathbb{R}^{p\times p}$ is invertible, and the metric d is well-known to be invariant to these, that is, $d\left(AX_1A', AX_2A'\right) = d(X_1, X_2)$ for all A. This means that the metric spatial depth in $(\mathcal{X}, d)$ already attains the "highest" form of invariance in $\mathcal{X}$ while still being computationally fast. Consequently, its invariance properties do not come with any kind of trade-off. Whereas, if d is instead taken to be the log-Euclidean metric $d(X_1, X_2) = \parallel Log(X_1) - Log(X_2) \parallel_F$, which is only invariant to orthogonal similarity transformations, then also the metric spatial depth achieves only this weaker form of invariance. In this case, an improvement can be obtained using the idea behind projection depths (Dyckerhoff, 2004): by defining the "projection" of X onto the direction $u \in \mathbb{S}^{p-1}$ as $X \mapsto u'Xu$, we can define the projection depth

$$D^*(\mu; P_X) := \inf_{u \in \mathbb{S}^{p-1}} D\left(u'\mu u; P_{u'Xu}\right),$$

where the depth D on the right-hand side is the metric spatial depth on the real line equipped with the metric $d(x_1, x_2) = |\log x_1 - \log x_2|$ and $P_{u'Xu}$ denotes the distribution of $u'Xu$. The resulting measure D^* is simultaneously robust and invariant to all similarity transforms, implying that the projecting gave us increased invariance at the cost of computational complexity.

ACKNOWLEDGMENTS

The author is grateful to the Associate Editor and an anonymous reviewer whose comments were of great help in improving the quality and presentation of the manuscript. This work

was supported by the Research Council of Finland (grants 347501 and 353769). Open access publishing facilitated by Turun yliopisto, as part of the Wiley - FinELib agreement.

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Kaggle at https://www.kaggle.com/datasets/zalando-research/fashionmnist.

ORCID

Joni Virta https://orcid.org/0000-0002-2150-2769

REFERENCES

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, *68*(3), 337–404.

Bachoc, F., Gonzalez-Sanz, A., Loubes, J.-M. & Yao, Y. (2024). *Wasserstein spatial depth*. arXiv preprint arXiv:2411.10646.

Bhatia, R. (2009). *Positive definite matrices*. Princeton University Press.

Bhattacharya, R., & Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. *Annals of Statistics*, *31*(1), 1–29.

Blom, G. (1976). Some properties of incomplete *U*-statistics. *Biometrika*, *63*(3), 573–580.

Chau, J., Ombao, H., & von Sachs, R. (2019). Intrinsic data depth for Hermitian positive definite matrices. *Journal of Computational and Graphical Statistics*, *28*(2), 427–439.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, *91*(434), 862–872.

Chen, M., Gao, C., & Ren, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Annals of Statistics*, *46*(5), 1932–1960.

Chen, Y., Dang, X., Peng, H. & Bart, H. L. (2008). Outlier detection with the kernelized spatial depth function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(2), 288–305.

Cholaquidis, A., Fraiman, R., Gamboa, F. & Moreno, L. (2023). Weighted lens depth: Some applications to supervised classification. *Canadian Journal of Statistics 51*(2), 652–673.

Conway, J. B. (1990). *A course in functional analysis* (Vol. *96*). Springer.

Dai, X., & Lopez-Pintado, S. (2023). Tukey's depth for object data. *Journal of the American Statistical Association*, *118*, 1760–1772.

De Micheaux, P. L., Mozharovskyi P. & Vimond, M. (2021). Depth for curve data and applications. *Journal of the American Statistical Association*, *116*(536), 1881–1897.

Dubey, P., & Müller, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika*, *106*(4), 803–821.

Dubey, P., & Müller, H.-G. (2022). Modeling time-varying random objects and dynamic networks. *Journal of the American Statistical Association*, *117*(540), 2252–2267.

Dyckerhoff, R. (2004). Data depths satisfying the projection property. *Allgemeines Statistisches Archiv*, *88*, 163–190.

Fang, K.-T., Kotz, S., & Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall.

Geenens, G., Nieto-Reyes, A., & Francisci, G. (2023). Statistical depth in abstract metric spaces. *Statistics and Computing*, *33*(2), 46.

Gijbels, I., & Nagy, S. (2017). On a general definition of depth for functional data. *Statistical Science*, *32*(4), 630–639.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. Wiley-Interscience.

Heinonen, L., Nyberg, H., & Virta, J. (2025). Weighted embedding and outlier detection of metric space data. *Advances in Data Analysis and Classification*, 1–37. https://link.springer.com/article/10.1007/s11634-025-00627-8#citeas

Kleindessner, M., & Von Luxburg, U. (2017). Lens depth function and k-relative neighborhood graph: Versatile tools for ordinal data analysis. *Journal of Machine Learning Research*, *18*(58), 1–52.
Lee, A. J. (1990). *U-statistics: Theory and practice*. Routledge.
Li, J., Cuesta-Albertos, J. A., & Liu, R. Y. (2012). DD-classifier: Nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association*, *107*(498), 737–753.
Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Annals of Statistics*, *18*, 405–414.
Liu, R. Y., & Singh, K. (1992). Ordering directional data: Concepts of data depth on circles and spheres. *The Annals of Statistics*, *20*(3), 1468–1484.
Liu, Z., & Modarres, R. (2011). Lens data depth and median. *Journal of Nonparametric Statistics*, *23*(4), 1063–1074.
Lyons, R. (2013). Distance covariance in metric spaces. *Annals of Probability*, *41*(5), 3284–3305.
Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. Wiley Online Library.
Mersmann, O. (2024). *microbenchmark: Accurate timing functions*. R package version 1.5.0. https://cran.r-project.org/web/packages/microbenchmark/index.html
Mosler, K., & Mozharovskyi, P. (2022). Choosing among notions of multivariate depth statistics. *Statistical Science*, *37*(3), 348–368.
Nieto-Reyes, A., & Battey, H. (2016). A topologically valid definition of depth for functional data. *Statistical Science*, *31*(1), 61–79.
Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, *1*(6), 327–332.
Paindaveine, D., & Van Bever, G. (2018). Halfspace depths for scatter, concentration and shape matrices. *Annals of Statistics*, *46*(6B), 3276–3307.
Pandolfo, G., Paindaveine, D., & Porzio, G. C. (2018). Distance-based depths for directional data. *Canadian Journal of Statistics*, *46*(4), 593–609.
Perlibakas, V. (2004). Distance measures for PCA-based face recognition. *Pattern Recognition Letters*, *25*(6), 711–724.
Rodrigues, É. O. (2018). Combining Minkowski and Chebyshev: New distance proposal and survey of distance metrics using k-nearest neighbours classifier. *Pattern Recognition Letters*, *110*, 66–71.
Schreurs, J., Vranckx, I., Hubert, M., Suykens, J. A. K., & Rousseeuw, P. J. (2021). Outlier detection in non-elliptical data by kernel MRCD. *Statistics and Computing*, *31*(5), 66.
Serfling, R. (2006). *Depth functions in nonparametric multivariate inference*. In *DIMACS series in discrete mathematics and theoretical computer science* (Vol. *72*, p. 1). American Mathematical Society.
Spiess, J. (1990). Some identities involving harmonic numbers. *Mathematics of Computation*, *55*(192), 839–863.
Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, *290*(5500), 2319–2323.
Tukey, J. W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians, Vancouver*, *1975*, 523–531.
Vardi, Y., & Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proceedings of the National Academy of Sciences*, *97*(4), 1423–1426.
Virta, J., Lee, K.-Y. & Li, L. (2022). Sliced inverse regression in metric spaces. *Statistica Sinica*, *32*, 2315–2337.
Xiao, H., Rasul, K. & Vollgraf, R. (2017). *Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms*. arXiv preprint arXiv:1708.07747.
Yang, M., & Modarres, R. (2018). β-skeleton depth functions and medians. *Communications in Statistics-Theory and Methods*, *47*(20), 5127–5143.
Zuo, Y., & Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics*, *28*, 461–482.

How to cite this article: Virta, J. (2026). Spatial depth for data in metric spaces. *Scandinavian Journal of Statistics*, 1–28. https://doi.org/10.1111/sjos.70054

APPENDIX

PROOFS

Proof of Theorem 1. Simplifying the squared reverse triangle inequality,

$$d^2(X_1, X_2) \geq |d(X_1, \mu) - d(X_2, \mu)|^2, \tag{A.1}$$

gives

$$d^2(X_1, \mu) + d^2(X_2, \mu) - d^2(X_1, X_2) \leq 2d(X_1, \mu)d(X_2, \mu).$$

Whereas, the squared triangle inequality,

$$d^2(X_1, X_2) \leq \{d(X_1, \mu) + d(X_2, \mu)\}^2, \tag{A.2}$$

gives

$$d^2(X_1, \mu) + d^2(X_2, \mu) - d^2(X_1, X_2) \geq -2d(X_1, \mu)d(X_2, \mu).$$

Hence,

$$| D(\mu; P) | \leq 1 + \frac{1}{2}\mathrm{E}\left\{ \frac{|d^2(X_1, \mu) + d^2(X_2, \mu) - d^2(X_1, X_2)|}{d(X_1, \mu)d(X_2, \mu)} \right\} \leq 2,$$

concluding the proof. ∎

Proof of Theorem 2. The claimed form for the influence function follows after observing that

$$\begin{aligned} &D(\mu; (1-\varepsilon)P + \varepsilon\delta_z) \\ &\quad = 1 - \frac{1}{2}\left[2(1-\varepsilon)^2\{1 - D(\mu; P)\} + 2\varepsilon(1-\varepsilon)\mathrm{E}\{h(X, z, \mu)\} + 2\varepsilon^2\mathbb{I}(\mu \neq z)\right], \end{aligned}$$

and that $| \mathrm{E}\{h(X, z, \mu)\} | \leq 2$ by the proof of Theorem 1. Furthermore, the same bound $| \mathrm{E}\{h(X, z, \mu)\} | \leq 2$, when combined with Theorem 3, yields the second part of the current theorem. ∎

Proof of Theorem 3. That $D(\mu; P)$ takes values in $[0, 2]$ follows directly from the proof of Theorem 1. For part (i), we observe that $D(\mu; P) = 0$ precisely when

$$\mathbb{I}(X_1 \neq \mu \cap X_2 \neq \mu)\frac{d^2(X_1, \mu) + d^2(X_2, \mu) - d^2(X_1, X_2)}{d(X_1, \mu)d(X_2, \mu)} = 2,$$

almost surely. By the definition of the indicator function and the proof of Theorem 1, this occurs if and only if $P(X = \mu) = 0$ and if equality is achieved in the squared reverse triangle inequality (A.1) almost surely. The latter of these is equivalent to the statement that $P(L[X_1, X_2, \mu] \cup L[X_2, X_1, \mu]) = 1$, proving claim (i).

Part (ii) is proven similarly, but by invoking the squared triangle inequality (A.2) instead of (A.1). ∎

Proof of Theorem 4. We begin by establishing some terminology. A sequence A_n of random variables in $\mathbb{R}$ is said to be a *D-sequence* (D is for "diverging") if $P(|A_n| \leq M) \to 0$ for every $M > 0$. Whereas, a random variable B taking values in $\mathbb{R}$ is said to be bounded if, for all $\varepsilon > 0$, there exists $K > 0$ such that $P(|B| \geq K) < \varepsilon$.

We next show that for any bounded random variable B and any D-sequence A_n, the quotient B/A_n converges to zero in probability. To see this, fix $\Delta, \varepsilon > 0$ and pick $K > 0$ such that $P(|B| \geq K) < \varepsilon$, and n_0 such that, for $n > n_0$, we have $P(|A_n| \leq K/\Delta) < \varepsilon$. Then,

$$\begin{aligned} &P(|A_n| \leq |B|/\Delta) \\ &\quad = P(|A_n| \leq |B|/\Delta \,|\, |B| \leq K) P(|B| \leq K) \\ &\qquad + P(|A_n| \leq |B|/\Delta \,|\, |B| > K) P(|B| > K). \end{aligned}$$

Now, $P(|A_n| \leq |B|/\Delta \,|\, |B| \leq K) \leq P(|A_n| \leq K/\Delta \,|\, |B| \leq K)$, allowing us to write

$$\begin{aligned} P(|A_n| \leq |B|/\Delta) &\leq P(|A_n| \leq K/\Delta) \\ &\quad + \{P(|A_n| \leq |B|/\Delta \,|\, |B| > K) - P(|A_n| \leq K/\Delta \,|\, |B| > K)\} P(|B| > K) \\ &\leq P(|A_n| \leq K/\Delta) + 2P(|B| > K) \\ &\leq 3\varepsilon, \end{aligned}$$

showing that $B/A_n = o_p(1)$.

Fix now an arbitrary point $\mu_0 \in \mathcal{X}$. It is clear that $d(X_1, \mu_0)$ is bounded. We next show that $d(X_1, \mu_n)$ is a D-sequence. Let $\varepsilon > 0$, $M > 0$ be arbitrary and let $K > 0$ be such that $P\{d(X_1, \mu_0) \geq K\} < \varepsilon$. Pick n_0 such that, for all $n > n_0$, we have $d(\mu_0, \mu_n) \geq M + K$.

Then, for $n > n_0$, we have, by the reverse triangle inequality, that

$$\begin{aligned} P\{d(X_1, \mu_n) \leq M\} &\leq P\{d(X_1, \mu_0) \geq d(\mu_0, \mu_n) - M\} \\ &\leq P\{d(X_1, \mu_0) \geq K\} \\ &\leq \varepsilon, \end{aligned}$$

showing that $d(X_1, \mu_n)$ is a D-sequence. Finally, it is obvious that $d(X_1, X_2)$ is bounded.

Consider now the quantity $R_{n4} := -d^2(X_1, X_2)/\{d(X_1, \mu_n) d(X_2, \mu_n)\}$. A straightforward computation reveals that the product of D-sequences is a D-sequence and, hence, we have that $R_{n4} = o_p(1)$. Next, by the triangle inequality, the quantity $R_{n2} := d(X_1, \mu_n)/d(X_2, \mu_n)$ satisfies

$$\left\{1 + \frac{d(X_1, X_2)}{d(X_1, \mu_n)}\right\}^{-1} \leq R_{n2} \leq 1 + \frac{d(X_1, X_2)}{d(X_2, \mu_n)},$$

showing that $R_{n2} = 1 + o_p(1)$. We can similarly show that $R_{n3} := d(X_1, \mu_n)/d(X_2, \mu_n) = 1 + o_p(1)$. Finally, it is straightforwardly proven that $R_{n1} := \mathbb{I}(X_1 \neq \mu_n \cap X_2 \neq \mu_n) =$

$1 + o_p(1)$. Consequently, denoting $R_n := R_{n1}(R_{n2} + R_{n3} + R_{n4}) = 2 + o_p(1)$, we have

$$D(\mu; P) = 1 - \frac{1}{2}\mathrm{E}(R_n).$$

As $| R_n |$ is uniformly bounded in n, we obtain the desired claim that $D(\mu; P) \to 0$. ∎

Proof of Theorem 5. Let $X_{1n}, X_{2n} \sim P_n$ be independent and let $X_1, X_2 \sim P$ be independent. Then $(X_{1n}, X_{2n}) \rightsquigarrow (X_1, X_2)$. Now, since the map $g : \mathcal{X}^2 \to \mathbb{R}$ defined as

$$g(x_1, x_2) := \mathbb{I}(x_1 \neq \mu \cap x_2 \neq \mu)\frac{d^2(x_1, \mu) + d^2(x_2, \mu) - d^2(x_1, x_2)}{d(x_1, \mu)d(x_2, \mu)},$$

is P-a.s. continuous, the continuous mapping theorem guarantees that

$$g(X_{1n}, X_{2n}) \rightsquigarrow g(X_1, X_2),$$

as $n \to \infty$ (note that g is measurable since it is P-a.s. equal to a continuous function). The function g is bounded by the proof of Theorem 1, implying that

$$\mathrm{E}\{g(X_{1n}, X_{2n})\} \to \mathrm{E}\{g(X_1, X_2)\},$$

and concluding the proof. ∎

Proof of Lemma 1. Plugging in the inner products, $d^2(a, b) = \langle a - b, a - b\rangle$, to the definition of $D(\mu; P)$, we obtain

$$D(\mu; P) = 1 - \mathrm{E}\left\{\left\langle \mathbb{I}(X_1 \neq \mu)\frac{X_1 - \mu}{\| X_1 - \mu \|}, \mathbb{I}(X_2 \neq \mu)\frac{X_2 - \mu}{\| X_2 - \mu \|}\right\rangle\right\}. \tag{A.3}$$

Now, for two independent random elements Y_1, Y_2 in $\mathcal{X}$ whose expected values exist, we have

$$\begin{aligned}\mathrm{E}(\langle Y_1, Y_2\rangle) &= \mathrm{E}\{\mathrm{E}(\langle Y_1 Y_2\rangle | Y_2)\}\\ &= \mathrm{E}\{\langle \mathrm{E}(Y_1), Y_2\rangle\}\\ &= \langle \mathrm{E}(Y_1), \mathrm{E}(Y_2)\rangle.\end{aligned}$$

Consequently, (A.3) takes the desired form

$$D(\mu; P) = 1 - \|\mathrm{E}\{\mathrm{sgn}(X - \mu)\}\|^2.$$

∎

Proof of Lemma 2. The depth $D(x_1; P)$ equals

$$1 - \frac{1}{2n^2}\sum_{i=2}^{n}\sum_{j=2}^{n}\frac{d^2(x_i, x_1) + d^2\left(x_j, x_1\right) - d^2\left(x_i, x_j\right)}{d(x_i, x_1)d\left(x_j, x_1\right)}.$$

Now, each index pair with $i = j$ is easily checked to contribute 2 to the sum, whereas each pair of distinct indices contributes the summand,

$$\frac{d^2(x_i, x_1) + d^2(x_j, x_1) - d^2(x_i, x_j)}{d(x_i, x_1)d(x_j, x_1)} \tag{A.4}$$

By the proof of Theorem 3, the quantity (A.4) is lower bounded by -2 and this bound is reached precisely if $d(x_i, x_j) = d(x_i, x_1) + d(x_1, x_j)$. The claim now follows. ∎

Proof of Lemma 3. The closed form for $D(1; P)$ is straightforwardly derived. For the lower bound, we observe that $s_2 := \sum_{i=1}^n p_i^2 \leq 1$ with equality if and only if all but one of the p_i equal zero. Consequently, $(1/2)(1 - s_2) + p_1 \geq 0$ with equality if and only if exactly one of $p_2, \ldots, p_n$ equals 1. For the upper bound, we first write $D(1; P) = 1 - (1/2)(1 - p_1)^2 - (1/2)(s_2 - p_1^2)$ from which it follows that $D(1; P) \leq 1$ with equality if and only if $p_1 = 1$. ∎

Proof of Theorem 7. We first compute the depth of an arbitrary point for a discrete, equispaced uniform distribution on $n = 2k$ points on the unit circle. We denote this distribution by P_n.

We label the $2k$ points such that our point of interest is 0 and the cyclical ordering of the points is $-(k-1), -(k-2), \ldots, -1, 0, 1, \ldots, (k-1), k$. That is, the point opposite of 0 on the cycle is k. Now, the expectation in the definition of $D(0; P_n)$ equals

$$\frac{1}{4k^2} \sum_i \sum_j \frac{d^2(i, 0) + d^2(j, 0) - d^2(i, j)}{d(i, 0)d(j, 0)}, \tag{A.5}$$

where the summations range over all points on the cycle but 0. To evaluate this sum, we first fix $i = 1, \ldots, k-1$ and observe that j can be chosen to lay in four different regions of the cycle, depending on whether the distances $d(j, 0)$ and $d(i, j)$ are realized clockwise or counterclockwise. Going individually through all four cases we observe that, for a fixed value of $i = 1, \ldots, k-1$, summing over j contributes the quantity

$$\frac{1}{4k^2} \left\{ 2i + \sum_{j=0}^{i} \frac{4ki + 4k(k-j) - 2i(k-j) - 4k^2}{i(k-j)} \right\}$$

to the expectation. By symmetry, each $i = -1, \ldots, -(k-1)$ contributes the same quantity as $|i|$ and, for $i = k$, the contribution is, using similar technique, seen to be $(2k-1)/(2k^2)$. Consequently, (A.5) takes (after simplification) the form

$$4 - \frac{2}{k} - \frac{3}{2k^2} + \frac{2}{k} H_{k-1} - 2 \sum_{i=1}^{k-1} \sum_{j=0}^{i} \frac{1}{i(k-j)},$$

where H_ℓ denotes the ℓth harmonic number. After observing that the double sum above equals $H_{k-1}/k + \sum_{j=1}^{k-1} (H_{k-1} - H_{k-1-j})/j$, the expression simplifies to

$$4 - \frac{2}{k} - \frac{3}{2k^2} - 2(H_{k-1})^2 + 2 \sum_{j=1}^{k-2} \frac{1}{k-1-j} H_j. \tag{A.6}$$

As an intermediate result, we next show that, for all $n \geq 1$,

$$\sum_{j=1}^{n} \frac{1}{n+1-j} H_j = (H_{n+1})^2 - \sum_{j=1}^{n+1} \frac{1}{j^2}. \tag{A.7}$$

To see this, we write

$$\begin{aligned}
\sum_{j=1}^{n} \frac{1}{n+1-j} H_j &= \sum_{j=1}^{n} \sum_{\ell=1}^{j} \frac{1}{\ell(j+1-\ell)} \\
&= \sum_{j=1}^{n} \sum_{\ell=1}^{j} \frac{1}{j+1} \left(\frac{1}{\ell} - \frac{1}{j+1-\ell} \right) \\
&= 2 \sum_{j=1}^{n} \frac{1}{j+1} H_j \\
&= 2 \sum_{j=1}^{n+1} \frac{1}{j} H_j - 2 \sum_{j=1}^{n+1} \frac{1}{j^2}.
\end{aligned}$$

The relation (A.7) now follows after observing that

$$2 \sum_{j=1}^{n+1} \frac{H_j}{j} = \sum_{j=1}^{n+1} \frac{1}{j^2} + (H_{n+1})^2,$$

by the example on page 850 in Spiess (1990). Plugging (A.7) into (A.6) now yields

$$D(0; P_n) = -1 + \frac{1}{k} - \frac{1}{4k^2} + \sum_{j=1}^{k} \frac{1}{j^2}.$$

By the classical Basel problem, as $n \to \infty$, the depth $D(0; P_n)$ approaches the value $\pi^2/6 - 1$. Hence, the proof is concluded as soon as we show that $P_n \rightsquigarrow P$ and invoke Theorem 5.

To see this, note that the weak convergence of distributions on $\mathbb{S}^1$ is equivalent to the pointwise convergence of the corresponding characteristic functions. That the characteristic function of P_n converges pointwise to that of P now follows by equations 3.5.14 and 3.5.16 in Mardia and Jupp (2000). ∎