

New approaches for analysing functional data - a focus on shape

Sami Helander

New approaches for analysing functional data - a focus on shape

Sami Helander

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall H304 of the school on 25 August 2022 at 12:00.

Aalto University
School of Science
Department of Mathematics and Systems Analysis

Supervising professor

Professor Pauliina Ilmonen, Aalto University School of Science, Finland

Thesis advisor

Professor Lauri Viitasaari, Uppsala University, Sweden

Preliminary examiners

Professor Thomas Verdebout, Université Libre de Bruxelles, Belgium

Adjunct Professor Sara Taskinen, University of Jyväskylä, Finland

Opponent

Professor Thomas Verdebout, Université Libre de Bruxelles, Belgium

Aalto University publication series

DOCTORAL THESES 102/2022

© 2022 Sami Helander

ISBN 978-952-64-0881-1 (printed)

ISBN 978-952-64-0882-8 (pdf)

ISSN 1799-4934 (printed)

ISSN 1799-4942 (pdf)

<http://urn.fi/URN:ISBN:978-952-64-0882-8>

Unigrafia Oy

Helsinki 2022

Finland



Printed matter
4041-0619

Author

Sami Helander

Name of the doctoral thesis

New approaches for analysing functional data - a focus on shape

Publisher School of Science**Unit** Department of Mathematics and Systems Analysis**Series** Aalto University publication series DOCTORAL THESES 102/2022**Field of research** Mathematics and Statistics**Manuscript submitted** 18 March 2022**Date of the defence** 25 August 2022**Permission for public defence granted (date)** 22 June 2022**Language** English **Monograph** **Article thesis** **Essay thesis****Abstract**

Functional data - sets of measurement sequences arising from a generating source of continuous nature - has become pervasive in many applications, across many fields of research. The commonly adopted methodology for exploring such data treats the observed units as functions, with continuous functional structure. This introduces a nigh boundless range of new modes of variability in shape and structure, unique to only this type of data. Thus, methodology encompassing the structural variability of functional data has risen to the attention in functional data literature.

In this thesis, we approach the shape features of functional data from three different angles, utilizing functional notions of statistical depth as well as metrics, sensitive to variations in structure. Our first approach demonstrates the advantages of data-driven methodology, that allows for the expert to utilize their contextual knowledge of the data. Thus, we provide a consistent framework, through statistical depth, for applying this ad hoc understanding to the analysis. In our second approach, we take a more universal outlook towards shape-encompassing methodology and introduce a new functional depth definition that utilizes some very recently discovered general notions of shape outlyingness. The third approach takes a step back from the context of depth and focuses on the more general concept of shape-sensitive metrics in functional spaces. We, in particular, introduce and study the properties of a new family of integrated metrics, that provide a notion of the overall local likeness (based on any pilot metric) of two curves.

Keywords functional data, shape-sensitivity, statistical depth, functional depth, integrated metric**ISBN (printed)** 978-952-64-0881-1**ISBN (pdf)** 978-952-64-0882-8**ISSN (printed)** 1799-4934**ISSN (pdf)** 1799-4942**Location of publisher** Helsinki**Location of printing** Helsinki **Year** 2022**Pages** 136**urn** <http://urn.fi/URN:ISBN:978-952-64-0882-8>

Tekijä
Sami Helander**Väitöskirjan nimi****Julkaisija** Perustieteiden korkeakoulu**Yksikkö** Matematiikan ja systeemianalyysin laitos**Sarja** Aalto University publication series DOCTORAL THESES 102/2022**Tutkimusala** Matematiikka ja Tilastotiede**Käsikirjoituksen pvm** 18.03.2022**Väitöspäivä** 25.08.2022**Väittelyluvan myöntämispäivä** 22.06.2022**Kieli** Englanti **Monografia** **Artikkeliväitöskirja** **Esseeväitöskirja****Tiivistelmä**

Funktionaalinen data - joukko jatkuvasta lähteestä kerättyjä mittaussarjoja - on yleistynyt useissa sovelluksissa, monilla eri tutkimuksen aloilla. Tällaista dataa tutkittaessa havaittuja yksikköjä käsitellään usein funktioina, joilla on jatkuva funktionaalinen rakenne. Tämä ilmenee lähes loputtomana skaalana uusia rakenteen ja muodon vaihteluja, joita esiintyy ainoastaan tällaisilla aineistoilla. Siispä, funktionaalisten aineistojen rakenteellisia vaihteluja huomioivat menetelmät ovat saaneet osakseen paljon huomiota funktionaalisen datan tutkimuksessa.

Tämä väitöskirja tarkastelee funktionaalisten aineistojen muoto-ominaisuuksia kolmella eri lähestymistavalla, hyödyntäen funktionaalista tilastollista syvyyttä sekä rakenteellisille vaihteluille herkkiä metriikoita. Ensimmäinen lähestymistapa havainnollistaa data-lähtöisen metodiikan hyötyjä, joissa asiantuntija voi hyödyntää aineistoon liittyvää kontekstuaalista tietämystään. Tarjoamme tilastolliseen syvyyteen pohjautuvan johdonmukaisen viitekehysten tilastolliselle analyysille, joka mahdollistaa ad hoc tietämyksen hyödyntämisen. Toisessa lähestymistavassamme tarkastelemme muoto-yksityiskohtia hyödyntävää metodiikkaa yleisemmästä näkökulmasta. Esittelemme uuden funktionaalisen tilastollisen syvyyden menetelmän, joka hyödyntää aivan hiljattain löydettyä muoto-poikkeavuuden käsitettä. Kolmas lähestymistapa ottaa askeleen kohti vieläkin yleisempää viitekehystä, ja tarkastelee muotoon sekä rakenteeseen liittyviä käsitteitä funktionaalisten metriikoiden kautta. Täsmällisemmin, esittelemme uuden integroitujen metriikoiden perheen, joka tarjoaa mittarin funktioiden kokonaisvaltaiselle paikalliselle samankaltaisuudelle.

Avainsanat funktionaalinen data, muoto-herkkyys, tilastollinen syvyys, funktionaalinen syvyys, integroitu metriikka**ISBN (painettu)** 978-952-64-0881-1**ISBN (pdf)** 978-952-64-0882-8**ISSN (painettu)** 1799-4934**ISSN (pdf)** 1799-4942**Julkaisupaikka** Helsinki**Painopaikka** Helsinki**Vuosi** 2022**Sivumäärä** 136**urn** <http://urn.fi/URN:ISBN:978-952-64-0882-8>

Preface

It is my great pleasure to thank the many people who have influenced me and supported me on this journey in academia, that I somewhat unexpectedly embarked on more or less as a result of inquiring about the possibility for a summer job.

First and foremost I wish to express my deepest gratitude to my supervisor Professor Pauliina Ilmonen and advisor Professor Lauri Viitasaari. I am extremely grateful for your outstanding guidance and support. You have always found the time to help me with any problem I might have had, mathematical or not, despite how big or small, and regardless of the time of the day – or in some occasions, night. Pauliina, not only are you a brilliant scholar, but you are also one of the sincerest, warmest and most open-minded people I know. I admire your constant enthusiasm towards mathematics, both in research and in teaching, and I hope to adopt some of it in my own future endeavours. Lauri, it has become very apparent that underneath your cool and laid-back demeanor, you share with Pauliina that same, almost child-like enthusiasm for technical mathematics and difficult problems. The depth and effortlessness of your mathematical insight never ceases to amaze me. I want to especially thank you for the patience and understanding you have had in guiding me through the ins and outs of many theoretical results that would have otherwise been way beyond me. Thank you both for the amazing mentorship and friendship throughout this journey. There still remains many intriguing problems and questions I hope to tackle together with you in the future.

I wish to thank the preliminary examiners Professor Thomas Verdebout and Adjunct professor Sara Taskinen for their careful examination of this thesis. Furthermore, I wish to thank Professor Verdebout for agreeing to act as my opponent.

I want to thank every one of the amazing people I have had the opportunity to work with during the completion of this thesis. Especially large thanks belongs to Professor Germain Van Bever and Professor Stanislav Nagy for your fruitful collaboration and patience in guiding me throughout our joint research endeavours. I deeply admire and respect you both. As

of writing this I feel like we are definitely overdue for our next research idea, what do you say?

I gratefully acknowledge the financial support from Emil Aaltosen Säätiö that has enabled me to focus on growing as a scholar throughout this journey. Thank you.

I wish to thank all of the people at Aalto University School of Science, Department of Mathematics and Systems Analysis and especially all of the current and former members of the Stochastics and Statistics research group, for these awesome past years. Special shout-outs to Matias Heikkilä for permanently embedding misuse of the word "faija" in my vocabulary and for some of the best office banter that even got us in trouble sometimes (for anyone reading, I swear he instigated it), and to Paavo Raittinen for all the blast we had both at office and during our evening sessions of our joint gaming passion. You know how they say that time flies in good company? Well it truly has flown.

I want to thank my family and friends, all of you loved ones, for your constant support throughout this journey and before it. There are too many of you great friends to list who have made a tremendous positive impact in my life. Nevertheless, I appreciate every one of you. I want to thank my sister Sarianna and my parents Jukka and Merja for your continuous love and support. I love and appreciate you tremendously. Jukka, especially you have been my greatest role-model and firmest support pillar to lean on in all the major decisions throughout my life. Words can't express how much I respect and value your influence in my life.

Finally, I would like to express my sincerest gratitude to my dearest Kaisa for all the laughter, joy and happiness you have brought me. You have become an integral part of my life and the focal point of each day. Thank you for laughing at my jokes, even at the ones that don't quite measure up to my usual sublime standard. Each day I'm becoming more and more convinced that, at least when it comes to humour, you and I collectively share a singular brain cell... I love you.

Helsinki, July 26, 2022,

Sami Helander

Contents

Preface	1
Contents	3
List of Publications	5
Author's contributions	7
1. Functional Data	9
1.1 Theoretical framework for functional data	12
1.1.1 Framework for Shape-sensitive metrics	14
1.2 Treatment of functional data in practice	15
2. Statistical Depth	23
2.1 Multivariate Depth	23
2.2 Functional Depth	29
3. Shape-sensitive metrics	39
3.1 Hausdorff distance	39
3.1.1 Integrated Hausdorff distance	40
3.2 Fréchet distance	42
3.2.1 Integrated Fréchet distance	44
4. Summaries of the articles	49
References	53
Publications	61

List of Publications

This thesis consists of an overview and of the following publications.

- I** S. Helander, G. Van Bever, S. Rantala, and P. Ilmonen. Pareto depth for functional data. *Statistics: A Journal of Theoretical and Applied Statistics*, vol. 54, no. 1, pp. 182–204, DOI: 10.1080/02331888.2019.1700418, February 2020.
- II** S. Nagy, S. Helander, G. Van Bever, L. Viitasaari, and P. Ilmonen. Flexible integrated functional depths. *Bernoulli*, vol. 27, no. 1, pp. 673–701, DOI: 10.3150/20-BEJ1254, February 2021.
- III** S. Helander, P. Laketa, P. Ilmonen, S. Nagy, G. Van Bever, and L. Viitasaari. Integrated shape-sensitive functional metrics. *Journal of Multivariate Analysis*, vol. 189, article 104880, DOI: 10.1016/j.jmva.2021.104880, May 2022.

Author's contributions

Publication I: “Pareto depth for functional data”

The original idea for this article originated from S. Rantala. The author developed the idea further and refined the presented approach, under the guidance of Prof. P. Ilmonen. The theoretical results were discovered by Prof. G. Van Bever and Prof. P. Ilmonen. The author implemented the method and designed the simulation study and the real-data analysis. The author contributed significantly to the writing of the paper and wrote the first version. All authors participated in the revision and polishing processes of the final version.

Publication II: “Flexible integrated functional depths”

The idea for this article originated from joint discussion between all authors and is based on previous work of Prof. S. Nagy. The theoretical results were discovered by Prof. S. Nagy. The author implemented the method and designed the real-data analysis under the guidance of Prof. S. Nagy and Prof. G. Van Bever. The author contributed significantly to the writing of the paper and wrote the first versions of the sections detailing the real-data analysis, while Prof. S. Nagy wrote the first versions of the sections detailing the theory. All authors participated in the revision and polishing processes of the final version.

Publication III: “Integrated shape-sensitive functional metrics”

The idea for this article originated from joint discussion between all authors. The approach taken in the paper was proposed by the author. Most of the theoretical results were discovered by the author, under the guidance

of Prof. L. Viitasaari. All authors contributed to the final formulation of the results. The methods were implemented by Prof. S. Nagy. The simulation study was designed by the author, under the guidance of Prof. S. Nagy and Prof. G. Van Bever. The author contributed significantly to the writing of the paper and wrote the first version. All authors participated in the revision and polishing processes of the final version.

1. Functional Data

With the development of modern measurement techniques and storage capacity, large and high-dimensional data sets have become commonplace. Such data sets are encountered across all fields of science. A common approach to handling such high-dimensional data is to assume that the observations are random functions, instead of random vectors. Thus, the observation sequences are seen as single entities, arising from sampling a process of continuous nature. The continuum over which the observation sequences are sampled can vary drastically, for example from spatial location or wavelength to frequency or concentration, etc. However, as time is the most prominently encountered such continuum, we henceforth refer to the continuum over which the observations are sampled as time t . By calling the data *functional*, we explicitly refer to the underlying continuous structure of the observed units; in principle, the measured processes could be observed at any arbitrary point in time, at any sampling frequency.

Due to the continuous nature of the sampled processes, the explicit functional form of the observations is unattainable. In practice, in its simplest (univariate) form, functional data consists of discrete sequences of value-index pairs (y_{ij}, t_{ij}) with $j = 1, 2, \dots, n_i$, where y_{ij} gives a snapshot of the value of the i th function at time $t_{ij} \in \mathcal{T}$ in the domain \mathcal{T} . Note that, in general, functional data can be multivariate with the components y_{ij} and t_{ij} being given by vectors of possibly differing lengths, describing for example high dimensional surfaces, high dimensional curves, or different measurements of the same observation instance that are to be analysed jointly. Furthermore, the observations in a set of functional data need not to be recorded over the same set of measurement points t_i . The number n_i of these measurement points, their placement, and their frequency over the domain can freely vary from observation to observation. This often discourages considering functional data simply as a set of high-dimensional vectors and prevents the direct use of multivariate methods on data of functional nature. Furthermore, the presence of noise in observation sequences, emerging from the measurement process, can limit the usefulness of raw functional data in direct analysis.

Thus, a common first step to analysing data of functional nature is to utilize the discrete measurement sequences in creating approximations of the underlying functions. This initial function recreation step may have a multitude of objectives on the resulting approximations, that we wish to fulfill at the same time. One objective is to create approximated curves that best match the known or assumed characteristics of the underlying process. These proxy curves allow us to approximate the value of the underlying continuous process at any arbitrary point $t \in \mathcal{T}$, even outside of the observed measurement grid $\{t_{ij}\}_{j=1}^{n_i}$. Often, when the underlying processes are known to be sufficiently smooth, we wish to create functional reconstructions that simultaneously allow us to approximate the derivative curves (of various degrees) of the underlying process, that can be used to aid the analysis. The goals of the functional recreation step may also include smoothing out the effects of measurement error or noise that is known to be present in the raw functional data. As this initial reconstruction step is a unique and important feature of functional data, it is discussed in more detail in Section 1.2.

One of the unique features of functional data is the immense amount of information carried by the continuous structure of the functions. The use of auxiliary curves such as derivatives of various degrees or time registration curves to support the analysis of functional data is a common approach. In many cases it is precisely the use of these auxiliary curves that reveal the important variance in functional data. For a classical example, consider the Berkeley growth data set from Tuddenham [1954], that has been prominently featured in functional data literature, and has been thoroughly analysed by numerous authors, including Ramsay et al. [1995], Ramsay and Li [1998], Ramsay and Silverman [2005], Jacques and Preda [2014] and Marron et al. [2015], just to name a few. The data depicts the growth curves of several boys and girls, measured over a set of 31 different ages. The growth curves of 15 randomly chosen girls are illustrated in Figure 1.1. From a visual inspection, the curves seem very similar and it is difficult to point out any characterizing features. However, the variability hidden in the data becomes apparent from the approximated growth acceleration curves (the second derivatives of the growth curves), that reveal a great deal of variance between the girls in the timing and intensity of the pubertal growth spurt. The growth acceleration curves of the 15 girls are depicted in Figure 1.2.

As data of functional nature has become pervasive in many applications across many different fields of science, development of methodology for analysing such data has become an important topic in the literature. Many of the originally multivariate statistical methods such as principal component analysis, canonical correlation analysis, regression modeling and statistical depth, have seen generalizations to the infinite-dimensional context of functional data (see for example Bosq [2000], Ramsay and Sil-

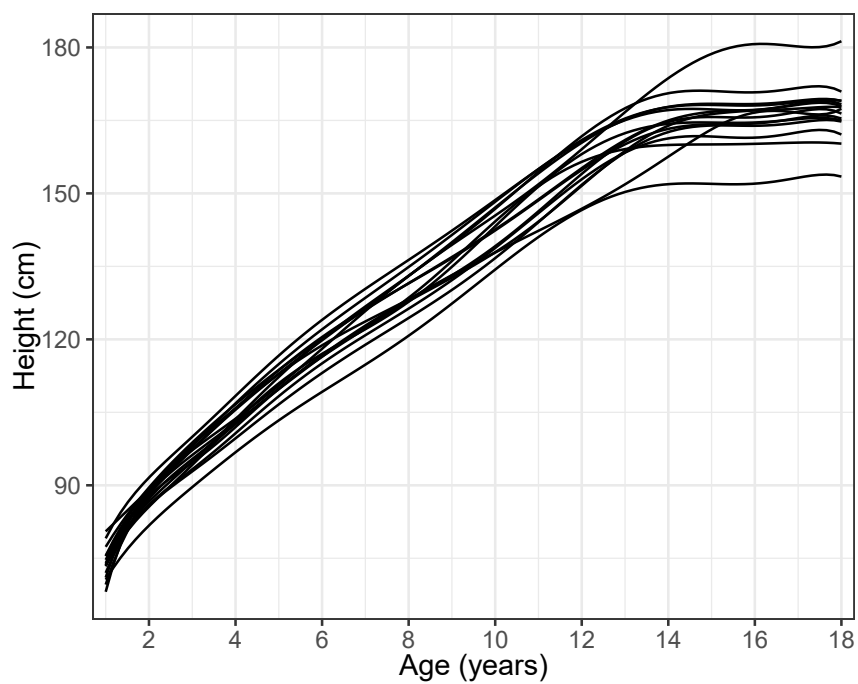


Figure 1.1. Growth curves of 15 randomly chosen girls from the Berkeley growth study.

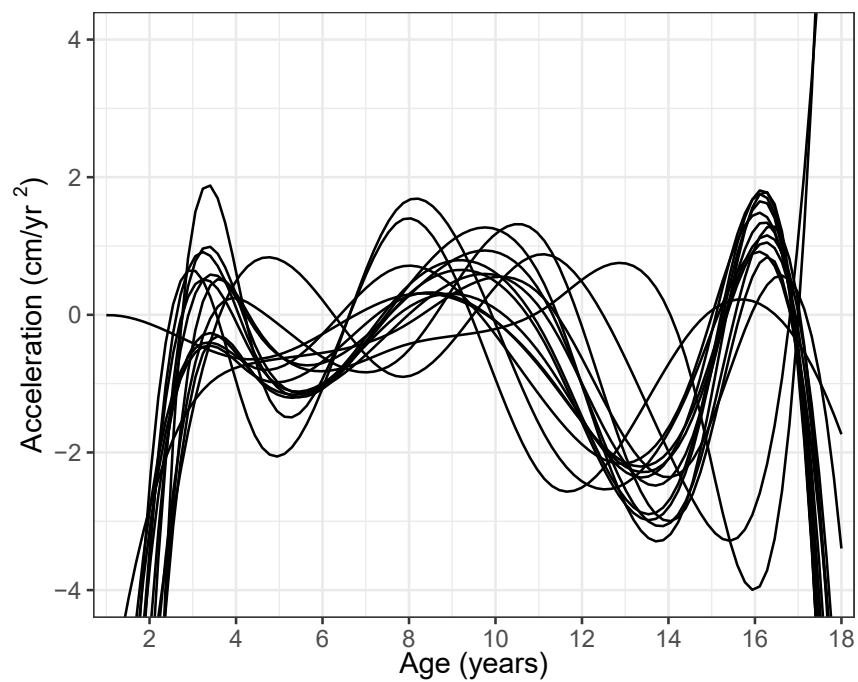


Figure 1.2. Growth acceleration curves of 15 randomly chosen girls from the Berkeley growth study

verman [2005], Ferraty and Vieu [2006], Horváth and Kokoszka [2012] and Hsing and Eubank [2015] as well as the discussion in Section 2.2).

However, stepping into the functional realm of infinite dimensionality brings with it a wide range of features in shape and structure, that need to be incorporated into the existing methodology. Indeed, shape variability and shape outlyingness have become integral topics in the FDA literature. Not only are many functional methods developed and extended specifically with considerations of shape variability in mind, but simply metrifying these concepts has become an important topic of discussion. This has turned out to be a difficult concept to approach, as measuring variance –or even outlyingness– in shape are both notions without a finite-dimensional analogue to expand upon. In \mathbb{R}^m , outlyingness is determined by the relative location of an observation (in some suitable metric) with respect to the distribution. While this approach can be directly applied to the functional context for example as the average pointwise (marginal) outlyingness of a function across the domain, these direct pointwise approaches are unable to detect the more subtle aspects of functional outlyingness such as centrally placed functions with structural differences. With this in mind, outlyingness in shape has received increasing attention in the literature, and there have been some recent advances for methods that consider the shape and structural variability of functional data. For more details, see the discussion in Section 2.2 as well as Epifanio [2008], López-Pintado and Romo [2009], Sun and Genton [2011], Claeskens et al. [2014], Nagy et al. [2017], Helander et al. [2020], Nagy et al. [2021], Helander et al. [2021] and Harris et al. [2021].

1.1 Theoretical framework for functional data

The theoretical framework commonly adopted in the functional data literature assumes that the observed functions are random realizations in a Hilbert space \mathcal{H} , defined on a compact support \mathcal{V} often taken to be a compact subset of \mathbb{R}^m , and equipped with a suitable inner product. Often, for univariate functions, the supporting compact set \mathcal{V} is without loss of generality taken to be the unit interval $[0, 1]$ as functions defined on any closed interval $I \subset \mathbb{R}$ can be identified with functions on $[0, 1]$ through a bijective linear mapping of the argument.

In fact, the observed functions are realizations of a *random function* $f : \Omega \times \mathcal{V} \rightarrow \mathcal{W}$, where (Ω, \mathcal{A}, P) denotes the underlying probability space. In practical considerations, the Ω is often omitted for notational simplicity. Thus, the random functions are associated with (randomly chosen) elements of the Hilbert space \mathcal{H} . When not specified, we refer to the Hilbert space of suitable \mathcal{W} valued functions $f : \mathcal{V} \rightarrow \mathcal{W}$ by $\mathcal{H}(\mathcal{V}, \mathcal{W})$, or simply by \mathcal{H} , when the context does not require the explicit examination of the sets \mathcal{V}

and \mathcal{W} . Imposing the functions with suitable regularity conditions guides the choice of the Hilbert space in question.

The most commonly encountered example of such space in the literature is

$$\mathcal{H} = L^2(\mathcal{V}, \mathbb{R}) := \left\{ f : \mathcal{V} \rightarrow \mathbb{R} ; \int_{\mathcal{V}} f(t)^2 \lambda(dt) < \infty \right\},$$

the set of real-valued square-integrable functions on a compact subset $\mathcal{V} \subset \mathbb{R}^m$ with respect to the Lebesgue measure λ , equipped with the inner product

$$\langle \cdot, \cdot \rangle_{L^2} : L^2(\mathcal{V}, \mathbb{R}) \times L^2(\mathcal{V}, \mathbb{R}) \rightarrow \mathbb{R} : (f, g) \mapsto \langle f, g \rangle_{L^2} := \int_{\mathcal{V}} f(t)g(t)\lambda(dt).$$

Another commonly studied space is the Sobolev space

$$W^{k,2}(I, \mathbb{R}) := \left\{ f : I \rightarrow \mathbb{R} ; \sum_{i=0}^k \int_I f^{(i)*}(t)^2 \lambda(dt) < \infty \right\},$$

the set of real-valued Lebesgue square-integrable functions on a closed interval $I \subset \mathbb{R}$, whose weak derivatives $f^{(i)*}$ up to order k are also Lebesgue square-integrable. Recall that $f \in L^2(I, \mathbb{R})$ is said to have a weak derivative $f^* \in L^2(I, \mathbb{R})$ if

$$\int_I f(t)\varphi'(t)dt = - \int_I f^*(t)\varphi(t)dt$$

holds for all infinitely differentiable functions $\varphi \in C^\infty(I)$, vanishing on the boundary of I . When equipped with the inner product

$$\langle \cdot, \cdot \rangle : W^{k,2}(I, \mathbb{R}) \times W^{k,2}(I, \mathbb{R}) \rightarrow \mathbb{R} : (f, g) \mapsto \langle f, g \rangle = \sum_{i=0}^k \langle f^{(i)*}, g^{(i)*} \rangle_{L^2},$$

the Sobolev space $W^{k,2}(I, \mathbb{R})$ also becomes a Hilbert space.

Further examples of more complex Hilbert spaces that are encountered in the literature include multivariate functional Hilbert spaces $\mathcal{H}(\mathcal{V}, \mathbb{R}^m)$ (Claeskens et al. [2014]), as well as Reproducing Kernel Hilbert spaces that are especially useful in the functional regression setting (Preda [2007], Hsing and Eubank [2015]). Aside from just Hilbert spaces, certain non-Hilbertian functional spaces are routinely considered. Examples of such spaces include: the space of continuous functions $C(\mathcal{V})$, commonly encountered in the literature as a simple example space, the space of k -times continuously differentiable functions $C^k(\mathcal{V})$ (Hsing and Eubank [2015]), and functions between (separable) metric spaces (Nieto-Reyes and Battey [2016], Helander et al. [2021]). However, in most cases, the functional spaces considered in the literature are assumed complete and equipped with an inner product (i.e. they are Hilbert spaces) to, at the very least, ensure approximations of the underlying functions in practice through projections on suitable elements (basis functions) in the considered functional space.

In the formal framework, the random functions f are seen as \mathcal{H} -valued random variables. That is, $f : \Omega \rightarrow \mathcal{H}$ are measurable mappings from the probability space (Ω, \mathcal{A}, P) to a measurable space $(\mathcal{H}, \mathcal{B})$, where \mathcal{B} is the σ -algebra generated by the open sets with respect to the norm induced by the inner product $\langle \cdot, \cdot \rangle$ on \mathcal{H} . In this sense, for $\mathcal{H}(\mathcal{V}, \mathcal{W})$ the random functions can be seen as functions $f(\omega, t)$ of two variables $t \in \mathcal{V}$ and $\omega \in \Omega$, such that, for any fixed t_0 , $f(\cdot, t_0) : \Omega \rightarrow \mathcal{W}$ is a random variable defined on (Ω, \mathcal{A}, P) while for any fixed ω_0 , $f(\omega_0, \cdot) : \mathcal{V} \rightarrow \mathcal{W}$ is a (fixed) element of $\mathcal{H}(\mathcal{V}, \mathcal{W})$. Thus, the elements of \mathcal{H} are in fact deterministic functions and the randomness arises from the underlying probability space (Ω, \mathcal{A}, P) .

In the theoretical framework of functional data, the sample of random functions f_1, \dots, f_n (i.i.d.) is assumed to be fully observed. That is, we have access to some known set of elements of \mathcal{H} that we can measure exactly at any arbitrary $t \in \mathcal{V}$. However, in practice functional data is only ever discretely observed. The treatment of the discrete measurement sequences in constructing continuous functional approximations is discussed in Section 1.2

1.1.1 Framework for Shape-sensitive metrics

In the following, we establish a simple theoretical framework for the discussion of Hausdorff and Fréchet distances in Sections 3.1 and 3.2. First, we recall the definition of a metric and a pseudometric.

Definition 1.1.1 (Metric). A metric on a set X is a function $d : X \times X \rightarrow [0, \infty)$ that satisfies the following conditions for all $x, y, z \in X$:

- (i) (*Non-negativity*): $d(x, y) \geq 0$
- (ii) (*Identifiability*): $d(x, y) = 0 \Leftrightarrow x = y$
- (iii) (*Symmetry*): $d(x, y) = d(y, x)$
- (iv) (*Triangle-inequality*): $d(x, y) \leq d(x, z) + d(z, y)$.

Additionally, d is said to be a *pseudometric* if it fulfils items (i), (iii), and (iv) of the above definition, but not item (ii).

Let $([0, 1], \lambda)$ be a measure space, where λ is the Lebesgue measure, and let $(\mathbb{R}^m, d_{\mathcal{V}})$ be a metric space with $d_{\mathcal{V}}$ Euclidean distance on \mathbb{R}^m . Furthermore, endow $([0, 1], \lambda)$ with the metric induced by λ . Consider the set of continuous functions

$$\mathcal{F} = \mathcal{C}([0, 1], \mathbb{R}^m) := \{f : [0, 1] \rightarrow \mathbb{R}^m ; f \text{ continuous}\}.$$

That is, functions f from the metric measure space $([0, 1], \lambda)$ to the metric

space (\mathbb{R}^m, d_V) , such that

$$\lim_{t \rightarrow s} d_V(f(t), f(s)) = 0.$$

The *curve* C_f , associated with the function f , is defined as

$$C_f : [0, 1] \rightarrow [0, 1] \times \mathbb{R}^m : t \mapsto (t, f(t)).$$

The *graph* G_f of f is then given by the image $G_f := C_f([0, 1]) \subset [0, 1] \times \mathbb{R}^m$. The classical definition of Hausdorff distance defines a metric between sets. On the other hand, the usual definition of Fréchet distance defines a metric between curves. Thus, Hausdorff and Fréchet distances between functions are naturally understood as the Hausdorff and Fréchet distances between the Graphs or the Curves (respectively) of the functions.

As the mappings $f \mapsto C_f$ and $C_f \mapsto G_f$ are bijective, the metrics between functions $f \in \mathcal{F}$, between curves C_f , and between graphs G_f , can all be interpreted interchangeably to produce metrics on suitable subsets of $[0, 1] \times \mathbb{R}^m$. We endow the space $[0, 1] \times \mathbb{R}^m$ with the usual q -metric:

$$d_C((t, v), (s, w)) := \begin{cases} (|t - s|^q + d_V(v, w))^{\frac{1}{q}} & \text{for } 1 \leq q < \infty, \\ \max\{|t - s|, d_V(v, w)\} & \text{for } q = \infty. \end{cases} \quad (1.1)$$

The natural choice $q = 2$ yields the metric induced by the L^2 norm on \mathbb{R}^{m+1} .

1.2 Treatment of functional data in practice

As functional data – data arising from measuring processes of continuous nature – is inherently only partially observable, often the first step to analysing such data is to recreate approximations of the observed functions that best match the known (or assumed) characteristics of the underlying process. These functional reconstructions allow us to approximate the value of the underlying functions at any arbitrary point t over the domain \mathcal{T} , enabling us to examine the observed functions over a unified measurement grid in cases where the individual sequences of measurement points $\{t_{ij}\}_{j=1}^{n_i}$ differ from an observation to another. In many cases, the reconstructed functions are also used to extract further information of the underlying processes, such as derivative curves (of various degrees) or time registration curves that can be leveraged in the analysis. In fact, in some cases, it is precisely the use of such auxiliary curves that allow us to access the key modes of shape and structural variation intrinsic to functional data.

A common approach to the functional reconstruction process is to approximate the underlying functions f by a linear expansion in terms of a suitable number K of known basis functions ϕ_k ,

$$f(t) \approx \sum_{k=1}^K c_k \phi_k(t),$$

where $c_k \in \mathbb{R}$, and the basis functions ϕ_k are chosen to be independent in an affine sense (see for example Ramsay and Silverman [2005], Hastie et al. [2009] and Horváth and Kokoszka [2012]). Furthermore, the chosen basis system must have the ability to approximate any function in the target class of functions arbitrarily well with a large enough number K of used basis function. Many of the commonly utilized basis systems can achieve satisfactory approximations of suitable classes of underlying functions with relatively small collections of basis functions. The coefficients c_k , for a given collection of basis functions $\{\phi_k\}$, can be approximated through regression techniques such as weighted least squares error minimization or roughness penalty error minimization. For a thorough overview on these and various other approximation techniques, see Ramsay and Silverman [2005] and Hastie et al. [2009].

The choice of a suitable basis has a crucial role in cases where approximation of the derivative curves is of interest. Due to the linear structure of the basis expansion, derivatives of the approximated functions directly translates to derivatives of the basis functions

$$f^{(m)}(t) = D^m f(t) \approx \sum_{k=1}^K c_k D^m \phi_k(t),$$

where for $m \geq 1$ we denote the differential operator $D^m f(t) = \frac{d^m}{dt^m} f(t)$.

The functional reconstruction step is also commonly called *smoothing* of the observation sequences, as, along with recreating approximations of the true functional forms of the observations, its secondary objective is to smooth out the effects of measurement error - noise - on the observed sequences. The presence of noise in the raw data is commonly modeled through an additive model, where the observed values are assumed to be a sum of the true value of the underlying process $f_i(t)$ together with an error term $\epsilon_i(t)$.

$$y_{ij} = f_i(t_{ij}) + \epsilon_i(t_{ij}).$$

The error terms $\epsilon_i(t)$ are often assumed to have some white noise properties, such as being independently distributed with mean zero and constant variance, and to be the major contributor to the observed roughness of the otherwise relatively smooth underlying process. Note that while smoothness assumptions of the underlying process are often leveraged in the reconstruction step to allow for non-parametric estimation of the approximating functions, the smoothness of the resulting approximations is not in itself the goal of the functional reconstruction process. Indeed, there are plenty of examples of very rough, even nowhere differentiable processes such as fractional Brownian motion, various Gaussian processes used in modeling, and processes governed by stochastic differential equations, that can in principle be analysed through the means of functional data analysis. Thus, smoothing of the observed sequences is in this context understood

relative to the roughness imposed on the observations by the noise in the measurement process.

It is worth noting that while the error process ϵ is routinely assumed to be white noise -like or stationary, these assumptions are often too simple for most functional data. For instance, it is common for the variance of the error to vary over the argument t , or for the errors to exhibit autocorrelations on varying time scales or cycles. This suggests that the ϵ , often deemed as erroneous noise, is itself a process we could model if we so choose. In practice, we often elect to simply ignore the effects of the error process and to focus on modeling the underlying functions f_i as including variable variance or autocorrelation structure of the residuals into the functional model can make model estimation difficult, escalate the computational load, and can still easily result in estimates that proved virtually no advantages over those achieved by assuming independence in the residuals. Nevertheless, in practice, it is up to the analyst's discretion to evaluate on a case-by-case basis the need for the considered functional model to also include modeling of the error process. In some cases, modeling the error process can be necessary to produce accurate functional models.

In the theoretical framework of functional data analysis, the functions are often assumed to be entirely observed, such that the true functional form of the underlying process giving rise to the observed random paths can be accessed. Thus, the underlying processes are often simply assumed to be continuous or to belong to another class of sufficiently well-behaved functions such as Lebesgue-integrable functions. However, in practice due to the presence of noise in the measurement process and functional data being only partially observable over a (relatively sparse) number of discrete measurement points, the continuity and structural assumptions are often supplemented with further smoothness assumptions to aid in constructing functional approximations from the raw data.

Commonly, two different approaches are encountered in literature. In the first approach, popularized by Ramsay and Silverman [2005], the underlying process itself is assumed smooth, leading to the construction of smooth approximations of the observed sample paths via a suitably chosen system of basis functions. In the second approach, often encountered in the context of functional principal components (see for example Yao et al. [2005], Hall et al. [2006] and Li and Hsing [2010]), the covariance function of the underlying process (instead of the observations) is assumed or constrained to be smooth. By utilizing Mercer's Theorem and Karhunen-Loève Theorem, this results in smooth approximations of the underlying process via the eigenfunctions of the covariance function. Other approaches to smoothing functional data include the use of kernels smoothers (see: Wand and Jones [1994]) and local linear or polynomial regression (see: Fan and Gijbels [1996]), to get rid of noisy variation in the raw observation sequences. These methods are commonly used as supplementary pre-

smoothing steps for especially noisy data, to allow for better functional approximations as well as to obtain estimates of the mean and covariance functions (see for instance Zhang and Wang [2016]).

In the first approach, it is directly assumed that the underlying process is smooth and has some suitable number of continuous derivatives. Consequently, the system of basis functions chosen for the function reconstruction is directly influenced by this assumption such that the resulting approximations exhibit the desired smoothness properties. Thus, we might attempt to simultaneously approximate the underlying functions together with a suitable number of derivative curves. Examples of such basis function systems that provide smooth functional approximations include the commonly used B-spline basis functions for smooth continuous functions with locally placed features (see: de Boor [2001]), the Fourier basis functions suited for periodic data (based on the common Fourier series) and the wavelet basis system for square-integrable functions that combine some of the features of splines and Fourier bases (see: Daubechies [1992]). For further examples of various other functional basis systems encountered in the literature, see Ramsay and Silverman [2005] and Hastie et al. [2009].

The B-spline basis functions of order m are defined over a fixed sequence of breakpoints τ as polynomials of order m that are positive over no more than m subsequent breakpoint intervals, join together smoothly at the breakpoint separating the non-zero regions of two B-spline functions, and possess continuous derivatives of order $m - 2$ that also coincide at these boundaries. Due to the B-spline basis functions having a compact support of fewer than m subsequent breakpoint intervals, they have similar computational advantages as orthogonal basis systems in that the computational complexity increases only linearly with respect to the number of basis functions used. Furthermore, although B-splines generally provide smooth and continuous functional approximations, they are capable of also approximating abrupt structural changes or even local discontinuity points by placing multiple congruent breakpoints at the same point in the domain.

Definition 1.2.1 (B-spline basis). Given a sequence of n breakpoints $\tau = t_0, t_1, \dots, t_n$ the k th B-spline B_k^m of order m over the breakpoint sequence is defined recursively by

$$B_k^1(t, \tau) := \begin{cases} 1, & \text{if } t_k < t < t_{k+1} \\ 0, & \text{otherwise,} \end{cases}$$

and

$$B_k^{m+1}(t, \tau) := \frac{t - t_k}{t_{k+m} - t_k} B_k^m(t, \tau) + \frac{t_{k+m+1} - t}{t_{k+m+1} - t_{k+1}} B_{k+1}^m(t, \tau).$$

Then, the B-spline basis functions ϕ_k of order m for the sequence of breakpoints τ are given simply by $\phi_k(t) = B_k^m(t, \tau)$.

The well known Fourier basis system suitable for approximating periodic data arises from the Fourier series:

$$f(t) = c_0 + c_1 \sin \omega t + c_2 \cos \omega t + c_3 \sin 2\omega t + c_4 \cos 2\omega t + \dots$$

where the parameter ω determines the period $2\pi/\omega$ of the series expansion. The Fourier basis system can prove to be especially useful in representing smooth functions that have stable behaviour and exhibit similar curvature throughout the domain. However, for functions with pronounced local features (such as abrupt changes in curvature that result in locally peaked paths for example) or potentially discontinuous paths or derivatives, the Fourier series may provide poor approximations.

Definition 1.2.2 (Fourier basis). Given a parameter ω , the Fourier basis functions φ_k are given by

$$\varphi_k(t) = \begin{cases} 1, & k = 0 \\ \sin r\omega t, & k = 2r - 1 \\ \cos r\omega t, & k = 2r. \end{cases}$$

The final functional basis example, the wavelet basis system, is suitable for approximating square-integrable functions on \mathbb{R} , especially when their derivative curves are of no particular interest. Informally, a wavelet is a wave-like function with an amplitude that starts from zero, increases or decreases oscillating around zero one or more times, and finally returns back to zero. The wavelet basis system consists of child wavelets $\psi_{jk}(t)$ that, for integers j and k , are formed as dilations and translations of a mother wavelet ψ of the form

$$\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k).$$

Within this basis system, for suitable J and K , the function $f(t)$ can then be approximated as

$$f(t) \approx \sum_{j=1}^J \sum_{k=1}^K c_{jk} \psi_{jk}(t),$$

for some real coefficients c_{jk} . The mother wavelet ψ is at the very least required to belong to the subspace $L^1 \cap L^2$ of functions that are both absolutely integrable as well as square-integrable, to ensure that the wavelet can be constrained to have zero mean and L^2 -norm one. The simplest possible example of a mother wavelet is the well known Haar mother wavelet introduced in Haar [1910]:

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & \text{otherwise.} \end{cases}$$

However, in practice, continuously differentiable functions with compact support are preferred, of which many examples exist.

Definition 1.2.3 (Wavelet basis). Given a mother wavelet $\psi \in L^1 \cap L^2$ that is absolutely integrable as well as square-integrable, for $j, k \in \mathbb{Z}$, the wavelet basis functions ψ_{jk} are given by

$$\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k).$$

In the second approach, commonly associated with the context of functional principal component analysis, the smoothness is assumed on the covariance process

$$K(s, t) = cov(f(s), f(t)),$$

instead of on the underlying processes themselves. Common methodology for estimating the covariance function utilizes local linear curve and surface smoothers (Fan and Gijbels [1996]) in estimating a smooth mean function and subsequently a smoothed covariance function from the raw functional data. For further details on the estimation process, see for example Yao et al. [2005], Hall et al. [2006], Li and Hsing [2010] and Zhang and Wang [2016].

By Mercer’s Theorem, for continuous $K(s, t)$, the set of eigenfunctions $\{\varphi_j\}_{j=1}^\infty$ of

$$\lambda_j \varphi_j(t) = \int_a^b K(s, t)\varphi_j(s)ds,$$

corresponding to a nonnegative sequence of eigenvalues $\{\lambda_j\}_{j=1}^\infty$, defines an orthonormal basis of $L^2([a, b], \mathbb{R})$. Furthermore, the eigenfunctions corresponding to the non-zero eigenvalues are continuous over $[a, b]$. The covariance function K admits to the representation

$$K(s, t) = \sum_{j=1}^\infty \lambda_j \varphi_j(s)\varphi_j(t).$$

Now, by utilizing the Karhunen-Loève theorem, random functions f on $L^2([a, b], \mathbb{R})$ with a mean function μ can be expressed as

$$f(t) = \mu(t) + \sum_{j=1}^\infty z_j \varphi_j(t),$$

where the random variables $z_j = \int_a^b (f(t) - \mu(t))\varphi_j(t)dt$ are uncorrelated and have zero-mean and variance λ_j . Thus, by estimating the eigenfunctions φ_j from a smoothed covariance process, or by utilizing a penalty on the roughness of the eigenfunction estimates, we obtain a system of (sufficiently) smooth basis functions in terms of which we can smoothly approximate square-integrable functions. Detailed discussion of smoothly approximating the eigenfunctions can be found in for example Rice and Silverman [1991] and Capra and Müller [1997].

While the smooth approximation of the underlying functions is a common approach in functional data analysis, it has been pointed out that

these smoothness assumptions exclude many rough processes from FDA and limit the scope of FDA to processes with smooth paths. As a remedy, Jouzdani and Panaretos [2021] propose a modified version of the usual covariance approximation procedure (discussed in Yao et al. [2005], Hall et al. [2006] and Li and Hsing [2010]), that enables approximation of the mean and covariance functions even under sparse and noisily observed settings, without relying on smoothness assumptions. The key observation of the approach is that rough but continuous sample curves may have a covariance function that is smooth outside of the diagonal $K(t, t)$. Thus, restricting the covariance approximation to the (lower) triangular region $((s, t) \in [a, b]^2 | a \leq t \leq s \leq b)$ and utilizing the symmetry over the diagonal, produces an approximation of the covariance function without requiring differentiability. Examples of processes encompassed by this approach, usually deemed too rough to be approached through the means of functional data analysis, include the standard Brownian motion, the standard Brownian bridge, the geometric Brownian motion, and the Ornstein-Uhlenbeck process.

2. Statistical Depth

In this section we explore the concept of statistical depth. First, we introduce the notion of statistical depth in its original multivariate context and discuss its applications. Then, we extend our understanding of depth to the functional context and explore the various new features that emerge from the infinite dimensional structure of functional data.

2.1 Multivariate Depth

Statistical depth is a nonparametric multivariate inferential tool that has received a lot of attention in the literature due to its several desirable, distributional feature revealing properties. In its original conception, the goal of statistical depth was to provide a natural center-outward ordering of multivariate data. That is, a depth function is any function $D : \mathbb{R}^m \rightarrow \mathbb{R} : x \mapsto D(x, P)$ that associates to each point $x \in \mathbb{R}^m$ a measure of its centrality with respect to the distribution P on \mathbb{R}^m . The values of the statistical depths usually range between 0 and 1, where the depth values close to 1 are attained by central points. This depth-based ordering of \mathbb{R}^m provides a natural basis for expanding the univariate rank and order statistics to the multivariate setting. Furthermore, statistical depth provides a useful tool in exploring distributional features of P , such as asymmetry, spread and shape, via nested depth-based contours. For further details, see for example Tukey [1975], Liu et al. [1999] and Serfling [2010]. As statistical depth coincides with the notion of quantiles in univariate settings, notions of multivariate quantiles based on depth regions or equi-depth contours have been studied in the literature. See for example Liu et al. [1999], Serfling [2010] and Hallin et al. [2010].

To establish a center-outward ordering of \mathbb{R}^m from a depth-based median outwards, we first require a relevant notion of center. If the underlying distribution P is symmetric, the concept of centrality is often tied to the center of symmetry (with respect to the corresponding notion of symmetry). Let $X \sim P$ be a random vector on \mathbb{R}^m . Commonly considered notions

of multivariate symmetry about $\theta_s \in \mathbb{R}^m$ include (i) *central symmetry*: $(X - \theta_s) \sim (\theta_s - X)$, (ii) *angular symmetry*: $(X - \theta_s)/\|X - \theta_s\|$ is centrally symmetric about the origin (Liu [1990]), and (iii) *halfspace symmetry*: $P(X \in H) \geq 1/2$ for every halfspace H containing θ_s (Zuo and Serfling [2000a]).

The notion of center outward ordering suggests that the points close to the center should be assigned the highest depth values, with the depth values decreasing as the distance to the depth-based center increases. Thus, in order to provide a unique ordering of \mathbb{R}^m and to preserve the intuition of center-outward ordering, conventional approaches to statistical depth definitions tend to ignore multimodality features of the underlying distribution P . However, this is a deliberate choice that was made in the general definition of statistical depth, that also comes with ramifications on the suitability of depth in analysis of distributions that might be multimodal or have non-convex support. While we shall return to this topic later, let us first take a look at some classical depth approaches as well as the axiomatic definition of statistical depth proposed by Zuo and Serfling [2000a].

Let $x \in \mathbb{R}^m$ an arbitrary point, the depth of which we might wish to measure. Let $P \in \mathcal{P}(\mathbb{R}^m)$, the set of all Borel probability measures on \mathbb{R}^m , and let $X \sim P$ an \mathbb{R}^m valued random variable. Tukey [1975] introduced perhaps the most universally applied and studied notion of depth, the *halfspace depth* HD , defined for a point $x \in \mathbb{R}^m$ with respect to a probability measure P as the minimum probability mass contained by any closed halfspace that also contains x . In the multivariate case $m > 1$, that is

$$HD(x, P) := \inf_{u \in \mathbb{R}^m, \|u\|=1} P(u^T(X - x) \geq 0),$$

where u belongs to the unit hypersphere on \mathbb{R}^m with respect to the usual Euclidean norm $\|\cdot\|$ on \mathbb{R}^m . Equivalently, in the univariate, case the halfspace depth is expressed as

$$HD(x, P) := \min\{P(X \leq x), 1 - \lim_{y \rightarrow x^-} P(X \leq y)\},$$

from which the connection of depth to univariate quantiles becomes immediately apparent.

Liu [1990] introduced *simplicial depth* SD , defined for $x \in \mathbb{R}^m$ with respect to P as the probability of x belonging to a random simplex in \mathbb{R}^m . That is,

$$SD(x, P) := P(x \in S[X_1, \dots, X_{m+1}]),$$

where X_1, \dots, X_{m+1} are i.i.d. random variables with distribution P , and $S[x_1, \dots, x_{m+1}]$ denotes the m -dimensional simplex – the set of all convex combinations of the vertices x_1, \dots, x_{m+1} on \mathbb{R}^m .

Liu and Singh [1993] considered these along with two more depth functions, the *Mahalanobis depth* and the *majority depth*, in the formulation of

a depth-based outlyingness measure. Mahalanobis depth MhD is defined for $x \in \mathbb{R}^m$ with respect to P through the inverse of the Mahalanobis distance of Mahalanobis [1936], from x to the distribution P . That is,

$$MhD(x, P) := \left(1 + (x - \mu_P)^T \Sigma_P^{-1} (x - \mu_P)\right)^{-1},$$

where μ_P and Σ_P denote the mean vector and covariance matrix of P respectively.

Singh [1991] considered the Majority depth MjD , defined for $x \in \mathbb{R}^m$ with respect to P as the probability of x being in a randomly chosen such halfspace of \mathbb{R}^m , that also contains at least $1/2$ the probability mass of P . That is,

$$MjD(x, P) := P(x \in H[X_1, \dots, X_m]),$$

where X_1, \dots, X_m are i.i.d. random variables with distribution P and $H[x_1, \dots, x_m]$ denotes the halfspace on the side of the boundary, given by the hyperplane containing the points x_1, \dots, x_m , with more probability mass.

Serfling [2002] introduced the *spatial depth* SpD , defined for $x \in \mathbb{R}^m$ with respect to P through the spatial quantiles of Chaudhuri [1996]. That is,

$$SpD := 1 - \left\| \mathbb{E} \left[\frac{x - X}{\|x - X\|} \right] \right\|,$$

where $\|\cdot\|$ denotes the usual Euclidean norm on \mathbb{R}^m and $\mathbb{E}[(x - X)/\|x - X\|]$ expresses the inverse of the spatial quantile function at x .

These and a plethora of other depth functions and their properties have been studied by various authors in the literature. Donoho and Gasko [1992] developed affine-equivariant multivariate location estimators based on the halfspace depth, that remain robust in high dimensions. Using halfspace depth, Yeh and Singh [1997] developed bootstrap confidence regions for multivariate descriptive parameters. Nolan [1992], Donoho and Gasko [1992] and Massé and Theodorescu [1994] (among others) studied the depth contours induced by halfspace depth. Liu and Singh [1993] developed a depth-based outlyingness measure, called the quality index, for assessing the outlyingness of a population (or a set) with respect to another. Rousseeuw and Ruts [1996], Ruts and Rousseeuw [1996] and Rousseeuw and Struyf [1998] developed algorithms for computing depth functions and depth based contours. Koshevoy and Mosler [1997] introduced a zonoid depth based on a zonoid trimming procedure. Rousseeuw and Hubert [1999] introduced the regression depth. Liu et al. [1999] developed several inferential and graphical tools for using depth in explorative data analysis, and considered examples using several depth definitions including the halfspace depth, the Mahalanobis depth, the convex hull peeling depth (Barnett [1976]), the Oja depth (based on the location measures introduced by Oja [1983] for centrally symmetric distributions), the simplicial depth, the majority depth, and the likelihood depth (Fraiman

and Meloche [1999]). Rousseeuw and Ruts [1999] provided insights to the computation of depth for population distributions and studied contour lines. Zuo and Serfling [2000b] studied depth-based multivariate scatter measures. Zuo and Serfling [2000a] categorized many of the previously proposed depth approaches to several depth classes, studied their properties, and proposed the first formal axiomatic approach to statistical depth. Several authors including Liu et al. [1999], Ghosh and Chaudhuri [2005b], Ghosh and Chaudhuri [2005a], Dutta and Ghosh [2012], Li et al. [2012] and Paindaveine and Van Bever [2015] studied depth-based approaches to supervised classification problems. Chen et al. [2009], Hlubinka et al. [2010], Agostinelli and Romanazzi [2011], and Paindaveine and Van Bever [2013] consider localized approaches to statistical depth that extend the usability of depth in statistical inference to multimodal and non-convexly supported distributions. In an extensive survey, Nagy et al. [2019] studied the relation of halfspace depth in statistics to some measures of symmetry considered in geometry and functional analysis. Motivated by these findings, Nagy and Dvořák [2021] studied the connection between the concept of illumination from convex geometry and the notion of statistical depth, and introduced the illumination depth for multivariate data.

Drawing upon the previous ideas of Liu [1990], the first formal axiomatic definition of statistical depth was formulated by Zuo and Serfling [2000a] as a bounded and non-negative function fulfilling the following four desirable properties: *affine invariance*, *maximality at center*, *monotonicity relative to the deepest point* and *vanishing at infinity*. These properties are collected together in the following definition.

Definition 2.1.1 (Zuo and Serfling [2000a]). Let \mathcal{P} denote the class of distributions on the Borel sets of \mathbb{R}^m and let $X \sim P_X$ be a \mathbb{R}^m valued random variable with $P_X \in \mathcal{P}$. The bounded and non-negative mapping $D(\cdot, \cdot) : \mathbb{R}^m \times \mathcal{P} \rightarrow \mathbb{R}$ is called a statistical depth function if it satisfies the following properties:

P1 (Affine invariance) $D(Ax + b, P_{AX+b}) = D(x, P_X)$ holds for any nonsingular matrix $A \in \mathbb{R}^{m \times m}$ and vector $b \in \mathbb{R}^m$.

P2 (Maximality at center) $D(\theta, P) = \sup_{x \in \mathbb{R}^m} D(x, P)$ holds for any $P \in \mathcal{P}$ having a unique center of symmetry θ with respect to some notion of symmetry.

P3 (Monotonicity relative to the deepest point) For any $P \in \mathcal{P}$ with a deepest point θ , it holds for all $\lambda \in [0, 1]$ that $D(x, P) \leq D(\theta + \lambda(x - \theta), P)$.

P4 (Vanishing at infinity) $D(x, P) \rightarrow 0$ as $\|x\| \rightarrow \infty$ holds for every $P \in \mathcal{P}$.

The corresponding sample depth, denoted by $D_n(x, P_n)$, is attained by

replacing P with its empirical measure P_n .

These properties collect precisely the requirements for a function to provide a meaningful center-outward ordering of points in \mathbb{R}^m relative to a distribution $P \in \mathcal{P}$. Property P1 ensures that the ordering provided by any depth function is unaffected by affine transformations of the space or the underlying measures, and truly considers the location of the points of \mathbb{R}^m in relation to the distribution P . For any distribution P with a uniquely defined center of symmetry, property P2 establishes this center of symmetry as the point of maximum depth with respect to which the points of \mathbb{R}^m are to be ordered. Property P3 maintains the center-outward ordering interpretation of depth, preventing the "inner points" of the distribution P from being assigned low depth values. Finally, P4 requires that the depth of a point should decrease towards zero as the distance from the center of the distribution increases without bounds.

In light of Definition 2.1.1, Zuo and Serfling [2000a] also studied a significant number of previous depth approaches, including depth extensions of some methods originally introduced in a non-depth context. Reportedly, many of the studied depth functions fail to satisfy some of the required properties P1–P4 and their use should therefore be carefully evaluated based on the application and context. Furthermore, drawing from ideas presented in the previous depth literature, Zuo and Serfling [2000a] identified several classes of general approaches to depth constructions and studied their properties. These classes were: (i) approaches to depth based on measuring the expected closeness of a point x to a random sample X_1, \dots, X_n of size n from P (example: simplicial depth, majority depth); (ii) approaches based on the inverse of the expected distance from x to the random sample X_1, \dots, X_n of size n from P (example: L^p depth¹); (iii) approaches based on the inverse of the outlyingness of x with respect to P (example: Mahalanobis depth); and (iv) approaches based on the minimum probability mass carried by a closed set $C \in \mathcal{C}$ containing x , in some suitable class of closed subsets \mathcal{C} of \mathbb{R}^m (example: halfspace depth).

However, as the goal of Definition 2.1.1 is to establish the criterion for a function to provide a meaningful center-outward ordering of \mathbb{R}^m , it becomes immediately apparent from examining the properties P1–P4 that such depth definitions are only suitable to describing the features of distributions that are unimodal and have convex support. Indeed, property P3 requires that the maximum depth with respect to any distribution $P \in \mathcal{P}(\mathbb{R}^m)$ is attained by at most a convex set of deepest points, with the depth values decreasing monotonously along any ray originating from any of the deepest points. Thus, depth is unimodal in the sense that the contour lines of the *depth regions* describe a single global peak maximizing

¹For $p > 0$, the L^p depth L^pD of $x \in \mathbb{R}^m$ with respect to P was defined by Zuo and Serfling [2000a] as $L^pD(x, P) = (1 + \mathbb{E}\|x - X\|_p)^{-1}$, where $X \sim P$ and $\|\cdot\|_p$ denotes the L^p -norm.

the depth value with no local maxima. Consider the depth regions $R(\alpha)$, introduced by Liu et al. [1999] as the set of points in \mathbb{R}^m attaining the depth value of at least α . That is, $R(\alpha) = \{x \in \mathbb{R}^m : D(x, P) \geq \alpha\}$, for some implicit distribution $P \in \mathcal{P}(\mathbb{R}^m)$. In suitable settings, these depth regions, and various closely related notions, have been established as useful tools in analysis of the features of the underlying distribution. However, while the depth regions for any $\alpha < \alpha_{\max} = \max_x D(x, P)$ are allowed to be non-convex, P3 requires them to still be fully connected and that for any $x \in R(\alpha)$, any $y \in R(\alpha_{\max})$ and any $\lambda \in [0, 1]$, it holds that $y + \lambda(x - y) \in R(\alpha)$, where $R(\alpha_{\max})$ denotes the region of maximum depth. This limits the usefulness of depth as defined in Definition 2.1.1 in cases where sensitivity to multimodality is essential, or the considered distributions might have convex supports. As an example, consider the widely applied halfspace depth that satisfies all of the properties P1–P4. As the halfspace depth is established to be quasi-concave for any positive measure on the Borel sets of \mathbb{R}^m (see: Rousseeuw and Ruts [1999]), it, in particular, is quasi-concave for any probability measure $P \in \mathcal{P}(\mathbb{R}^m)$, meaning that it induces contour lines that are unimodal (in the above sense) and convex throughout, regardless of the underlying distribution.

To remedy these issues, there have been attempts in the literature at extending the applicability of statistical depth to multimodal and non-convexly supported distributions, while still retaining the desirable properties P1–P4 when considered with unimodal and convexly supported distributions. These extensions are called *local depths*, as they aim to bridge the gap between the global ordering of \mathbb{R}^m and sensitivity to the local features of the distribution. Chen et al. [2009] introduced the kernelized spatial depth, that achieves localization through a suitably chosen kernel function, by computing the spatial depth in a feature space induced by the chosen kernel. Hlubinka et al. [2010] considered a weighted halfspace depth, based on considering weighted probabilities in the halfspaces rather than the probability of the halfspace. Agostinelli and Romanazzi [2011] introduced localized versions of the halfspace depth and the simplicial depth, defined through replacing the halfspaces with finite-width slabs and by restricting the maximum volume of the considered simplices, respectively. Paindaveine and Van Bever [2013] introduced perhaps the most universally applicable local depth construction, that can be used in conjunction with any statistical depth function to provide a measure of local centrality at any level of locality. To base the construction purely on depth, the localization is achieved through the use of the depth-based neighbourhoods of Paindaveine and Van Bever [2015].

2.2 Functional Depth

As extremely high dimensional data has become prevalent across many fields of science, extending methods of statistical inference for such data has received a lot of attention in the literature. Due to its richness, parametric modeling of functional data that accurately captures its diverse features is known to be difficult. Thus, due to its nonparametric nature and versatility in exploration of the distributional characteristics of multivariate data, it is not surprising that statistical depth has been seen as a potentially attractive tool in the functional context as well.

In the following, let $x \in \mathcal{H}(\mathcal{V})$ for some suitable Hilbert space \mathcal{H} with a compact support \mathcal{V} . Let $P \in \mathcal{P}(\mathcal{H})$, the space of all probability measures on \mathcal{H} , let $X \sim P$ be a \mathcal{H} valued random variable, and let $P_{X(t)}$ denote the marginal distribution of X at time $t \in \mathcal{V}$. For simplicity, let $\mathcal{V} \subset \mathbb{R}$.

The first instance of a functional depth was introduced by Fraiman and Muniz [2001], who proposed an integrated approach to depth by measuring the (often univariate) point-wise depth of x at t , $D(x(t), P_{X(t)})$, across the domain, and then taking the integral. That is, for a suitable statistical depth D , the corresponding *integrated functional depth* FD of $x \in \mathcal{H}$ with respect to P is given by

$$FD(x, P) := \int_{t \in \mathcal{V}} D(x(t), P_{X(t)}) dt.$$

Following this idea, a wide variety of definitions and approaches to functional depth have been introduced in the literature. Cuevas et al. [2007] proposed the random projection depth and the h -mode depth. The random projection depth of x considers the univariate depths of the projections $\langle u, x \rangle$, when projected on random elements $u \in \mathcal{H}$ through $\langle u, x \rangle = \int_{\mathcal{V}} u(t)x(t)dt$, with respect to the projected distribution, $P_{\langle u, X \rangle}$. To gain a representative value for depth, an average is taken. In a double random projection version of the method, the bivariate sample of random projections of x and its derivative curve x' were considered. In the h -mode depth, the depth of $x \in \mathcal{H}$ with respect to P is determined as the expectation of the kernel-weighted distance (in a norm suitable to \mathcal{H}) from x to $X \sim P$. A fixed tuning parameter h is used in re-scaling of the chosen kernel. Cuesta-Albertos and Nieto-Reyes [2008] considered the random Tukey depth for functional data, a version of the random projection depth where the averaging is replaced by taking a minimum over a set of randomly chosen projections. Cuevas and Fraiman [2009] considered the integrated dual depth, determined for x with respect to P through the expected depth of $f(x)$ with respect to $P_{f(X)}$, where $f \in \mathcal{H}^*$ are operators in the dual space \mathcal{H}^* . The expectation is computed with respect to a suitable distribution on \mathcal{H}^* . In our setting with \mathcal{H} a Hilbert space, the operators f are simply inner products with fixed elements $u \in \mathcal{H}$. In the

paper, Cuevas and Fraiman [2009] also considered Banach valued functions. López-Pintado and Romo [2009] introduced the band depth and the modified band depth. For $x \in \mathcal{H}$ with respect to P , the band depths consider the belonging of the graph $G_x = \{(t, x(t)) : t \in \mathcal{V}\}$ to the bands $B(X_1, \dots, X_j) = \{(t, y) : t \in \mathcal{V}, y \in [\min_{k=1, \dots, j} X_k(t), \max_{k=1, \dots, j} X_k(t)]\}$ over different values of j , where X_1, \dots, X_j are i.i.d. random functions from the distribution P . The band depth is given by the sum of the probabilities of G_x being entirely contained within the band $B(X_1, \dots, X_j)$ over a range of values $j = 2, \dots, J$. Instead of the probabilities, the more flexible modified version considers the expected proportions of the domain over which x is contained in such band of j random functions, and a sum is taken over the range of values $j = 2, \dots, J$. López-Pintado and Romo [2011] introduced the the half region depth and the modified half region depth. For $x \in \mathcal{H}$ with respect to P , the half region depths consider the belonging of G_x to the half regions $A(X) = \{(t, a) : t \in \mathcal{V}, a \geq X(t)\}$ and $B(X) = \{(t, b) : t \in \mathcal{V}, b \leq X(t)\}$, for some $X \sim P$. The half region depth is given by the minimum of the probabilities of G_x belonging entirely to either $A(X)$ or $B(X)$. For the modified version, instead of the probabilities, the minimum of the expected proportions of the domain over which G_x belongs to $A(X)$ or $B(X)$ is considered. Mosler [2013] proposed the idea of considering an infimum of the pointwise depths over the domain, instead of an integral. Furthermore, Mosler [2013] introduced the Φ -depths, a class of infimum-based depths with a close relation to the integrated dual depth. For Φ -depth, for some suitable set $\Phi \subset \mathcal{H}^*$ in the dual space, the univariate depths of $f(x)$ with respect to $P_{f(X)}$ are considered over all $f \in \Phi$, and the depth value is taken to be the infimum. Claeskens et al. [2014] introduced the integrated multivariate functional depth, a direct extension of the integrated functional depth FD . In the related approach, the univariate function $x \in \mathcal{H}$, for which the depth is evaluated, is supplemented at each time point $t \in \mathcal{V}$ with additional information, arising for example from its derivative or time-registration curves, to form a J -dimensional vector. Then, the multivariate depth of the supplemented vector is evaluated with respect to the joint distribution of the marginal distributions of the supplementary curves (and x , of course). That is, let $x^J = (x(t), x_1^*(t), \dots, x_j^*(t))^T$ denote the joint vector of the values of x and its J supplementary curves x_1^*, \dots, x_j^* at time t . Then, the depth of $x \in \mathcal{H}$ with respect to $P \in \mathcal{P}(\mathcal{H})$ is attained by integrating the multivariate depth of x^J with respect to $P_{(X(t), X_1^*(t), \dots, X_j^*(t))^T} \in \mathcal{P}(\mathbb{R}^{J+1})$ over the domain, where $X \sim P$ and X_1^*, \dots, X_j^* are supplemental curves derived from X . In applications, the use of an additional weight function (that integrates to one) allows to control the sensitivity of the depth to the functional variations over different regions of the domain. Chakraborty and Chaudhuri [2014a] as well as Chakraborty and Chaudhuri [2014b] studied the functional spatial depth, a straightforward extension of the multivariate spatial depth

S_pD to the functional context by replacing the Euclidean norm in \mathbb{R}^m with an appropriate norm in \mathcal{H} . Sguera et al. [2014] and Sguera et al. [2016] studied the kernelized functional spatial depth, a generalized version of the kernelized spatial depth of Chen et al. [2009] that naturally extends to the functional setting in a way similar to the functional spatial depth. Nagy et al. [2017] introduced the classes of J th order integrated and infimal depths, shape sensitive approaches to functional depth that expand upon the integrated functional depth of Fraiman and Muniz [2001] and its corresponding infimal approach. For these classes of functional depths, for a fixed J , the multivariate depth of the vector $(x(t_1), \dots, x(t_J))^T$ of multiple time margins t_1, \dots, t_J is considered with respect to the joint distribution $P_{(X(t_1), \dots, X(t_J))^T} \in \mathcal{P}(\mathbb{R}^J)$, where $X \sim P$. Then, either an integral or an infimum is taken over all combinations of $(t_1, \dots, t_J)^T \in \mathcal{V}^J$, where \mathcal{V}^J is the J -fold cartesian product of \mathcal{V} . Helander et al. [2020] proposed a depth-based application-driven inferential method for functional data called Pareto-depth. In the approach, the observed curves $x \in \mathcal{H}$ are mapped to a J -variate vector through several mappings $f_j : \mathcal{H} \rightarrow \mathbb{R}$, $j = 1, \dots, J$, that measure some characteristics or features of the observations that are deemed of importance in the applied context. Then, the depth of the curve x with respect to P is determined as the multivariate Pareto-depth of its associated vector with respect to the joint distribution of the mappings, $P_{(f_1(X), \dots, f_J(X))^T} \in \mathcal{P}(\mathbb{R}^J)$, where $X \sim P$. Nagy et al. [2021] considered the class of J th order k th moment integrated depths, and studied their properties. The approach expands upon the J th order integrated depths of Nagy et al. [2017] by involving the depth distribution of the cross-sectional half-space depth in to the functional depth consideration. Harris et al. [2021] introduced the elastic depths, based on separate analysis of the amplitude and phase variability of the functions. First, outlyingness measures are defined based on both the amplitude distance (difference in the function value) and phase distance (difference in the timing of the structural features) between the functions. The elastic depths are then defined through the inverse of the outlyingness measures.

However, despite the rich body of literature that has developed around functional statistical depth and its uses in functional data analysis, functional depth still lacks a formal, agreed upon definition and a set of criteria to fulfill. This problem has been discussed by several authors, including Nieto-Reyes [2011], Nagy et al. [2016], Nieto-Reyes and Battey [2016] and Gijbels and Nagy [2017], all of whom have made attempts at bringing together a formal set of criterion for functional depth. Nieto-Reyes and Battey [2016] proposed the first attempt at defining such a set of formal properties for functional depth, collected under Definition 2.2.2. Furthermore, Nieto-Reyes and Battey [2016] provide an extensive analysis of the most prominent functional depth examples in terms of the proposed properties.

Definition 2.2.1 (Convex hull (Nieto-Reyes and Battey [2016])). Let $(\mathcal{H}, \mathcal{A}, P)$ be a probability space where \mathcal{H} is a Hilbert space with compact support \mathcal{V} , \mathcal{A} is the σ -algebra on \mathcal{H} generated by the open d -metric balls for some suitable metric d , and $P \in \mathcal{P}(\mathcal{H})$, the space of all probability measures on \mathcal{H} . Define \mathcal{E} to be the smallest set in the σ -algebra \mathcal{A} such that $P(\mathcal{E}) = P(\mathcal{H})$. Then the convex hull of \mathcal{H} with respect to P is defined as

$$CH(\mathcal{H}, P) := \{x \in \mathcal{H} : x(v) = \alpha L(v) + (1 - \alpha)U(v), v \in \mathcal{V}, \alpha \in [0, 1]\},$$

where $L := \{\inf_{x \in \mathcal{E}} x(v) : v \in \mathcal{V}\}$ and $U := \{\sup_{x \in \mathcal{E}} x(v) : v \in \mathcal{V}\}$.

Definition 2.2.2 (Nieto-Reyes and Battey [2016]). Let $(\mathcal{H}, \mathcal{A}, P)$ be a probability space as in Definition 2.2.1. The bounded and non-negative mapping $D(\cdot, \cdot) : \mathcal{H} \times \mathcal{P} \rightarrow \mathbb{R}$ is called a statistical functional depth if it satisfies the following properties:

FP1 (Distance invariance) $D(f(x), P_{f(X)}) = D(x, P_X)$ holds for any $x \in \mathcal{H}$ and $f : \mathcal{H} \rightarrow \mathcal{H}$ such that for any $y \in \mathcal{H}$ we have $d(f(x), f(y)) = a_f d(x, y)$, where $a_f \in \mathbb{R} \setminus \{0\}$.

FP2 (Maximality at center) For any $P \in \mathcal{P}(\mathcal{H})$ with a unique center of symmetry $\theta \in \mathcal{H}$ with respect to some notion of functional symmetry, it holds that $D(\theta, P) = \sup_{x \in \mathcal{H}} D(x, P)$.

FP3 (Strictly decreasing with respect to the deepest point) For any $P \in \mathcal{P}(\mathcal{H})$ for which $D(z, P) = \max_{x \in \mathcal{H}} D(x, P)$ exists, $D(x, P) < D(y, P) < D(z, P)$ holds for any $x, y \in \mathcal{H}$ such that $\min\{d(y, z), d(y, x)\} > 0$ and $\max\{d(y, z), d(y, x)\} < d(x, z)$.

FP4 (Upper semi-continuity in x) $D(x, P)$ is upper semi-continuous as a function of x ; i.e. for all $x \in \mathcal{H}$ and all $\epsilon > 0$ there exists a $\delta > 0$ such that $\sup_{y : d(x, y) < \delta} D(y, P) \leq D(x, P) + \epsilon$.

FP5 (Receptivity to convex hull width across the domain) It holds that $D(x, P_X) < D(f(x), P_{f(X)})$ for any $x \in CH(\mathcal{H}, P)$ for which $D(x, P) < \sup_{y \in \mathcal{H}} D(y, P)$ and $f : \mathcal{H} \rightarrow \mathcal{H}$ such that $f(y(v)) = \alpha(v)y(v)$ where $\alpha(v) \in (0, 1)$ for all $v \in L_\delta$ and $\alpha(v) = 1$ for all $v \in L_\delta^C$, where

$$L_\delta := \arg \sup_{\mathcal{H} \subset \mathcal{V}} \left\{ \sup_{x, y \in \mathcal{C}(\mathcal{H}, P)} d(x(\mathcal{H}), y(\mathcal{H})) \leq \delta \right\},$$

for any $\delta \in [\inf_{v \in \mathcal{V}} d(L(v), U(v)), d(L, U))$ such that $\lambda(L_\delta) > 0$ and $\lambda(L_\delta^C) > 0$.

FP6 (Continuity in P) For all $x \in \mathcal{H}$, for all $P \in \mathcal{P}(\mathcal{H})$ and for every $\epsilon > 0$, there exists a $\delta > 0$ such that $|D(x, Q) - D(x, P)| < \epsilon$ holds P -almost surely for all $Q \in \mathcal{P}(\mathcal{H})$ for which $d_{\mathcal{P}}(Q, P) < \delta$ P -almost surely, where $d_{\mathcal{P}}$ metricises the topology of weak convergence.

As there is no unique concept of symmetry for a distribution of functions $P \in \mathcal{P}(\mathcal{H})$, Nieto-Reyes and Battey [2016] propose the following alternative property as a replacement:

FP2G (Maximality at Gaussian process mean) For P a zero-mean, stationary, almost surely continuous Gaussian process on \mathcal{V} , it holds that $D(\theta, P) = \sup_{x \in \mathcal{H}} D(x, P) \neq \inf_{x \in \mathcal{H}} D(x, P)$, where θ is the zero mean function.

In the paper, Nieto-Reyes and Battey [2016] provide the following intuitions for the proposed properties. The first three of the properties are proposed as straightforward analogues of the corresponding multivariate properties. Property FP1 requires for the functional depth to remain invariant under such transformations from \mathcal{H} to \mathcal{H} , that preserve the relative distances between elements of \mathcal{H} (in the metric d on \mathcal{H}) up to some scaling factor. In particular, FP1 ensures that functional depths remain unaffected by re-centering of the space around some function $\theta \in \mathcal{H}$.

Together, properties FP2 and FP3 lead to functional depth defining a center-outward ordering of \mathcal{H} in the spirit of the multivariate definition. The property FP2 requires that for any $P \in \mathcal{P}$ with a uniquely defined center of symmetry $\theta \in \mathcal{H}$ (with respect to a suitable definition of functional symmetry), the notions of (maximum) depth and symmetry coincide at θ . The alternative property FP2G extends to the functional context the notion in the multivariate settings that for such distributions $P \in \mathcal{P}(\mathbb{R}^m)$ for which multiple notions of symmetry coincide at $\theta \in \mathbb{R}^m$, this point should also attain the maximum depth value. As the most common example of such distribution is the Gaussian distribution, for which the expected value, the median and many notions of symmetry all coincide, the straightforward extension of this notion to the functional context is stated with respect to the functional counterpart of the Gaussian distribution - the Gaussian process.

To achieve a center-outward ordering of \mathcal{H} , the elements of \mathcal{H} should be prescribed lower depth values as the distance from the depth-based center of P increases. Thus, property FP3 requires that the depth values assigned to each of the functions are inversely related to the size of the d -metric ball around the deepest point in which they belong to. This further implies that as the distance (in metric d) of a point from the depth-based center $z \in \mathcal{H}$ increases, the assigned depth values tend towards the infimum of the depth in the distribution P ; $\lim_{x:d(x,z) \rightarrow \infty} D(x, P) = \inf_{x \in \mathcal{H}} D(x, P)$. In

the functional case, however, this infimum of the depth is not required to be 0.

Property FP4 aims to preserve the ability of depth functions of revealing features of the underlying distribution. In the multivariate setting, statistical depth has been linked with a notion of multivariate quantiles and has been shown to reveal properties of the underlying distribution. Thus, in order for functional depth to similarly reveal features of the underlying functional distribution, it should satisfy similar continuity properties as a cumulative distribution function.

The goal of property FP5 is to reduce the possible effects of measurement error in the functional observations on the resulting depth-based ordering of the data. In some applications, the functional values exhibit little variability and overlap significantly over some particular subset $L \subset \mathcal{V}$ of the domain. It is argued that the depth should thus heed the functional values over $\mathcal{V} \setminus L$ to a far greater extent than the values over L where, due to the lack of variance and the frequent overlapping of the observations, the ordering of the functions is of less relevance. As a particular example for integrated depth approaches for achieving this property, Nieto-Reyes and Battey [2016] propose the use of an additional weight function in the integration, that allows placing less weight on the depth values over such regions L .

Property FP6 seeks to achieve two essential goals. The first of the goals is to allow the use of depth in statistical inference, by ensuring that the sample depth, based on an empirical distribution P_n , converges almost surely to its population counterpart. That is, $D(\cdot, P_n) \rightarrow D(\cdot, P)$ almost surely for all sequences P_n such that $P_n \rightarrow P$ as $n \rightarrow \infty$. The secondary goal is to address the subtle fact that due to the inherent partial observability of functional data, even the empirical distribution P_n is inaccessible (in its entirety) and needs to be approximated. An empirical distribution P_n on \mathcal{H} would consist of a collection of point masses at $X_1, \dots, X_n \in \mathcal{H}$. However, in practice, the functional observations are obtained as sequences of discrete measurements, from which an empirical distribution of reconstructed observations \hat{P}_n , approximating P_n , is constructed through the use of functional approximation techniques (see Section 1.2). Thus, FP6 ensures the validity of our statistical inference even when based on such empirical distributions of reconstructed observations, given that the reconstruction process yields $\hat{P}_n \rightarrow P_n$ almost surely (as the measurement frequency of the individual observations, and thus the number of discrete points over which the function values are recorded, increases). Furthermore, the property FP6 plays a role in the robustness of functional depth in regards to the existence of outlying observations in the data.

Based on these notions, Gijbels and Nagy [2017] provide a detailed analysis of the proposed properties FP1-FP6 and their implications. In particular, it is pointed out that while many of the proposed properties

make intuitive sense, Definition 2.2.2 is not yet complete as a general definition for functional depth, and great care needs to be taken in extending multivariate concepts to functional spaces. Below, we summarize the findings of Gijbels and Nagy [2017] and the suggested alternatives to the proposed properties FP1-FP6. For more details, see the thorough analysis in Gijbels and Nagy [2017].

In the paper, it is found that the proposed property FP1 is very demanding in functional spaces. While it might be justified to pursue such invariance with respect to mappings that are multiples of an isometry, property FP1 is in fact not a direct generalization of the multivariate property P1. As a substitute, Gijbels and Nagy [2017] propose two alternative properties; invariance with respect to scalar-affine mappings², or invariance with respect to function-affine mappings³, both of which have been previously considered in the literature (see for instance López-Pintado and Romo [2009] and Claeskens et al. [2014]).

For the property FP2, Gijbels and Nagy [2017] provide further insights into the extensions of the multivariate concepts of central symmetry and halfspace symmetry into the functional context. In support of the intuition of FP2, two additional alternative properties are suggested; maximality at the center with respect to functional central symmetry, and maximality at the center with respect to functional halfspace symmetry.

The formulation of the proposed property FP3 is found to be restrictive in functional settings, and it is in fact much stronger than the multivariate property P3. It is reported that none of the commonly considered examples of functional depth satisfy this property. Thus, this property is proposed to be amended to the following straightforward functional extension of P3:

FP3D (Decreasing with respect to the deepest point) For any $P \in \mathcal{P}(\mathcal{H})$ such that $D(z, P) = \sup_{x \in \mathcal{H}} D(x, P)$, it holds that $D(z, P) > \inf_{x \in \mathcal{H}} D(x, P)$ and $D(x, P) \leq D(z + \lambda(x - z), P)$ for all $\lambda \in [0, 1]$ and $x \in \mathcal{H}$.

Furthermore, it was found that the property of vanishing at infinity (multivariate P4) is not necessarily guaranteed by FP3 or FP3D alone, and should be established separately.

While the intuition behind property FP5 is sensible in some applications where the presence of noise in the measurements is of concern, it might not be well suited for a –general– definition of functional depth; Gijbels and Nagy [2017] found that FP5 has negative implications on the arguably more generally desirable properties of invariance with respect to function-

²A mapping $f : \mathcal{F} \rightarrow \mathcal{F}$ is called *scalar-affine*, if $f(x) = ax + b$, where \mathcal{F} is a suitable functional space, $x, b \in \mathcal{F}$ and $a \in \mathbb{R} \setminus \{0\}$.

³A mapping $f : \mathcal{F} \rightarrow \mathcal{F}$ is called *function-affine*, if $f(x) = ax + b$, where $\mathcal{F}(\mathcal{V})$ is a suitable functional space defined over the compact set \mathcal{V} , $x, a, b \in \mathcal{F}$, $ax \in \mathcal{F}$, and $a(t) \neq 0$ for all $t \in \mathcal{V}$. The available functional spaces \mathcal{F} are restricted among those, for which the operation $ax \in \mathcal{F}$ makes sense.

affine transformations or scalar-affine transformations, proposed as an amendment to the property FP1.

Finally, it is agreed that a property such as FP6, ensuring the validity of statistical inference based on empirical distributions P_n , is of extreme importance. However, Gijbels and Nagy [2017] argue that the formulation of FP6 is perhaps too weak, and stronger versions may be considered. In practice, the depth values assigned to particular observations are usually not the focus of interest, and it is more important that the whole surface of depth values is well approximated by the finite sample version. Thus, Gijbels and Nagy [2017] propose an alternative version of FP6 utilizing the stronger, uniform continuity in P instead:

FP6U (Uniform continuity in P) For every $\epsilon > 0$, there exists $\delta > 0$ such that for any $P, Q \in \mathcal{P}(\mathcal{H})$ for which $d_{\mathcal{P}}(P, Q) < \delta$, it holds that $\sup_{x \in \mathcal{H}} |D(x, P) - D(x, Q)| < \epsilon$, where $d_{\mathcal{P}}$ metricises the topology of weak convergence in $\mathcal{P}(\mathcal{H})$.

The notions concerning shape and shape outlyingness of functions have recently risen to the attention in the FDA literature. In particular, recent contributions to functional depth have begun emphasizing sensitivity to shape outlyingness as a desirable property for a general notion of functional depth. See for instance the discussion in López-Pintado and Romo [2009], Sun and Genton [2011], Claeskens et al. [2014], Nagy et al. [2017], Helander et al. [2020], Nagy et al. [2021] and Harris et al. [2021]. Many of the early approaches to functional depth focus on the pointwise centrality of $x \in \mathcal{H}$ in P as a measure of its depth, for example through integrating (or considering the infimum of) a pointwise measure of centrality (such as multivariate depth) over the domain. As a result, such approaches often ignore the more general structural features of x which can lead to centrally placed shape-outliers – functions whose pointwise values $x(t)$ reside close to the center of $P_{X(t)}$ over (most) of the margins $t \in \mathcal{V}$, yet who differ from the other functions of P in their behaviour – being given high depth values. Examples of such central shape-outliers are readily available in the literature; see for example the simulated examples in Nagy et al. [2017], Helander et al. [2020], Harris et al. [2021] and Helander et al. [2021].

Although sensitivity to outlyingness in shape is a widely recognized problem in functional depth literature, there have been remarkably few attempts at formalizing the phenomenon outside of conceptual examples. Perhaps the first formal definition of shape outlyingness was given by Nagy et al. [2017]. This definition is built recursively as follows; Assume that some notion of multivariate outlyingness –such as multivariate statistical depth– is agreed upon. Let $x \in \mathcal{H}(\mathcal{V})$, $P \in \mathcal{P}(\mathcal{H})$ and $X \sim P$. If there exists any $t \in \mathcal{V}$ such that $x(t) \in \mathbb{R}$ is outlying with respect to the distribution

$P_{X(t)} \in \mathcal{P}(\mathbb{R})$, then x is considered a first-order outlier with respect to P . Such functions exhibit at least a single (local) feature that is deemed outlying with respect to values usually attained by $X \sim P$ at that point in the domain. For $J = 2, 3, \dots$, x is called a J th order outlier with respect to P , if x is not a j th order outlier for any $j < J$, and there exists a set of points $(t_1, \dots, t_J)^T \in \mathcal{V}^J$ such that $(x(t_1), \dots, x(t_J))^T \in \mathbb{R}^J$ is outlying with respect to the distribution $P_{(X(t_1), \dots, X(t_J))^T} \in \mathcal{P}(\mathbb{R}^J)$. For $J \geq 2$, J th order outlyingness of x is related to a difference in shape compared to the functions from P ; 2nd order outlyingness signifies a difference in linear growth – the 1st derivative, 3rd order outlyingness signifies a difference in convexity or concavity – the 2nd derivative, and so on. Based on this definition of outlyingness, Nagy et al. [2017] introduce the shape-sensitive classes of J th order integrated and infimal depths, upon which the J th order k th moment integrated depths of Nagy et al. [2021] further expand. Other examples of functional depth approaches, that explicitly emphasise shape sensitivity in their construction, include the multivariate functional halfspace depth of Claeskens et al. [2014], the functional Pareto depth of Helander et al. [2021] and the elastic depth of Harris et al. [2021].

3. Shape-sensitive metrics

Quantifying the similarity or dissimilarity in shape between functions has attracted concurrent interest on many fields of research. For example, in machine learning and computer vision, various shape and pattern matching applications have become commonplace. A commonly used approach to such problems is to employ the Hausdorff or Fréchet distances between the graphs or the curves of the functions. See Huttenlocher et al. [1993], Rucklidge [1997], Yi and Camps [1999], Veltkamp and Hagedoorn [2001], De Carvalho et al. [2006], Alt and Godau [1992], Alt et al. [2003], Brakatsoulas et al. [2005], Aronov et al. [2006] and Jiang et al. [2008], and the discussion provided in Sections 3.1 and 3.2.

However, both of these distance definitions are based on a global supremum, limiting the analysis to the most drastic feature difference between the functions, instead of considering their overall likeness. Thus, in the literature, some attention has been devoted to developing integrated versions of these metrics. In the following, we explore the concepts of the Hausdorff and Fréchet distances, as well as various approaches to defining integrated versions of these distances, that have been previously considered in the literature. We highlight some of the difficulties that are encountered in attempting to define these integrated alternatives, and discuss the caveats that the previously considered approaches might have when applied to the context of functional data.

3.1 Hausdorff distance

The Hausdorff distance d_H has been established as a widely useful tool in places where the concept of the proximity of different sets to one another is crucial. In the case of functions, the Hausdorff distance d_H between two functions $f, g \in \mathcal{F}$ can be naturally understood as the Hausdorff distance between their graphs $G_f, G_g \subset [0, 1] \times \mathbb{R}^m$.

Definition 3.1.1 (Hausdorff distance). Consider the functions $f \in \mathcal{F}$ and $g \in \mathcal{F}$. The Hausdorff distance d_H between the functions is defined as the

Hausdorff distance between their graphs:

$$d_H(f, g) := \max \left\{ \sup_{t \in [0,1]} \inf_{s \in [0,1]} d_C(C_f(t), C_g(s)), \sup_{s \in [0,1]} \inf_{t \in [0,1]} d_C(C_f(t), C_g(s)) \right\}. \quad (3.1)$$

This definition can be intuitively understood as a ball rolling along the graph of either of the functions. Then, the Hausdorff distance between the graphs is given as the radius of smallest such ball that, while rolling along one of the graphs, paints every point on the other graph. Thus, Hausdorff distance can be equivalently defined through

$$d_H(f, g) = \inf \{ \epsilon \geq 0 : G_f \subset G_g^\epsilon \text{ and } G_g \subset G_f^\epsilon \},$$

where

$$G^\epsilon := \bigcup_{x \in G} \{ z \in [0, 1] \times \mathbb{R}^m : d_C(z, x) \leq \epsilon \}.$$

In any metric space, the Hausdorff distance defines a proper metric for the set of all non-empty compact subsets of the space, while still remaining as a pseudometric even if the subsets are not compact. Since graphs G_f of continuous functions $f \in \mathcal{F}$ with a compact support $[0, 1]$ are compact sets on $[0, 1] \times \mathbb{R}^m$, the Hausdorff distance defines a metric for the graphs. This, in turn, can be interpreted as a metric for the functions.

The Hausdorff distance is readily used in a wide variety of different research fields. Among others, d_H has been applied in sample path analysis for planar Brownian motion (Taylor [1964]), image matching, pattern recognition and machine vision (Huttenlocher et al. [1993], Rucklidge [1997], Yi and Camps [1999], Veltkamp and Hagedoorn [2001], De Carvalho et al. [2006], to name just a few), the approximation of the global attractors in dynamical systems (Dellnitz and Hohmann [1997], Aulbach et al. [2005]), in fractal geometry (Falconer [2004]), and in convergence analysis to the Pareto set in multi-objective optimization problems (Schütze [2004], Dellnitz et al. [2005], Schütze et al. [2008]).

3.1.1 Integrated Hausdorff distance

The Hausdorff distance is known to severely punish single outliers, which can be a considerable drawback in some applications. For example, in evaluation of the performance of stochastic search methods, known to generate some outliers, the Hausdorff distance is of limited use.

As such, there have been some efforts in the literature in developing an integrated version of the Hausdorff metric. Baddeley [1992] developed one such integrated distance in both the discrete and the general case (the Baddeley's delta metric), and studied its use in image processing. In the proposed approach, an alternative (yet equivalent) representation of the usual Hausdorff distance is studied, in which the supremum is

simply replaced with an integral. Marron and Tsybakov [1995] took a more straightforward approach, by directly replacing the *sup inf* notion of pointwise distances between sets with an integral. This approach was applied as an error criterion in a nonparametric curve estimation setting. A similar idea, albeit in a finite setting, was also considered by Schütze et al. [2012], who proposed the p -averaged Hausdorff distance d_{aH}^p by replacing the *sup inf* notion with an L_p mean.

Definition 3.1.2 (p -averaged Hausdorff (Schütze et al. [2012])). Let X and Y be two non-empty finite subsets of a metric space $(\mathcal{M}, d_{\mathcal{M}})$. For a given $p \geq 1$, define the p -averaged Hausdorff distance d_{aH}^p as

$$d_{aH}^p(X, Y) := \max \left\{ \left(\frac{1}{|X|} \sum_{x \in X} D(x, Y)^p \right)^{\frac{1}{p}}, \left(\frac{1}{|Y|} \sum_{y \in Y} D(y, X)^p \right)^{\frac{1}{p}} \right\},$$

where $|X|$ denotes the cardinality of the set X (respectively for Y) and $D(x, Y)$ is some $d_{\mathcal{M}}$ -based notion of distance from the point $x \in X$ to the set Y (respectively for $y \in Y$ and X).

For $p = \infty$, the p -averaged Hausdorff distance d_{aH}^{∞} is interpreted as the usual Hausdorff distance d_H .

However, compared to the usual Hausdorff distance, d_{aH}^p is only defined for finite sets and it only yields an inframetric¹ instead of a metric.

Vargas and Bogoya [2018] introduced a modification to d_{aH}^p called the (p, q) -averaged Hausdorff distance $d_{aH}^{p,q}$, that generalizes the p -averaged Hausdorff distance and more importantly, is a proper metric (for finite sets) for $1 \leq p, q < \infty$.

Definition 3.1.3 ((p, q) -averaged Hausdorff (Vargas and Bogoya [2018])). Let X and Y be two non-empty finite subsets of a metric space $(\mathcal{M}, d_{\mathcal{M}})$. For given $p, q \in \mathbb{R} \setminus \{0\}$, the (p, q) -averaged Hausdorff distance $d_{aH}^{p,q}$ is defined as

$$d_{aH}^{p,q}(X, Y) := \max \{ \mathbf{GD}^{p,q}(X, Y \setminus X), \mathbf{GD}^{p,q}(Y, X \setminus Y) \},$$

where

$$\mathbf{GD}^{p,q}(X, Y) = \left(\frac{1}{|X|} \sum_{x \in X} \left(\frac{1}{|Y|} \sum_{y \in Y} d_{\mathcal{M}}(x, y)^q \right)^{\frac{p}{q}} \right)^{\frac{1}{p}}.$$

¹Given a parameter $h \geq 1$, the h -inframetric on a set X is a function $d_h : X \times X \rightarrow [0, \infty)$ that, for all $x, y, z \in X$, satisfies the properties (i)–(iii) of Definition 1.1.1 together with the following weaker version of triangle inequality:

$$(iv)' : d_h(x, y) \leq h \max\{d(x, z), d(z, y)\}.$$

Here, $d_{aH}^{p,q}$ generalizes d_{aH}^p in the sense that for disjoint subsets it holds that $\lim_{q \rightarrow -\infty} d_{aH}^{p,q} = d_{aH}^p$.

Bogoya et al. [2018] further develop the idea for arbitrary measurable sets by showing that the power means in $\text{GD}^{p,q}$ can be replaced with their integrated counterparts. Let (Σ, d_Σ, μ) be a metric measure space. Let $\mathfrak{M}(\Sigma)$ be the σ -algebra of all measurable subsets of Σ , and let $\mathfrak{M}_{<\infty}(\Sigma)$ denote those elements of $\mathfrak{M}(\Sigma)$ with finite measure.

Definition 3.1.4 ((p, q) -averaged Hausdorff for measurable sets (Bogoya et al. [2018])). Let $X, Y \in \mathfrak{M}_{<\infty}(\Sigma)$. For given $p, q \in \mathbb{R} \setminus \{0\}$, define the (p, q) -averaged Hausdorff distance $d_{aH}^{p,q}$ as

$$d_{aH}^{p,q}(X, Y) := \max \{ \text{GD}^{p,q}(X, Y \setminus X), \text{GD}^{p,q}(Y, X \setminus Y) \},$$

where

$$\text{GD}^{p,q}(X, Y) = \left(\frac{1}{\mu(X)} \int_X \left(\frac{1}{\mu(Y)} \int_Y d_\Sigma(x, y)^q d\mu(y) \right)^{\frac{p}{q}} d\mu(x) \right)^{\frac{1}{p}},$$

and where X and Y are assumed to be disjoint when $p < 0$ or $q < 0$.

Bogoya et al. [2018] show that this definition gives a semimetric² on the collection $\mathfrak{M}_{<\infty}(\Sigma)$ of all measurable finite subsets of Σ , and is a proper metric for $1 \leq p, q < \infty$ if the sets are disjoint.

3.2 Fréchet distance

The Fréchet distance is another well-established tool often used as a shape similarity measure for curves in shape matching tasks. Whereas the Hausdorff distance can give a good indication of the proximity of two point sets, it may not always be a suitable choice for graphs or curves arising from functions. Indeed, examples of curves that may be perceived as dissimilar, yet have a small Hausdorff distance, are readily available.

Consider $f(t) = \cos(kt)$ and $g(t) = \cos(kt + \pi)$, defined on $t \in [0, 2\pi]$. Then, $d_H(f, g) = \frac{\pi}{k}$ which becomes arbitrarily small as k increases. Yet, the two curves are perceived to be rather dissimilar, always being on opposite phases with opposite signs of derivatives. In such cases, the Fréchet distance is a more suitable distance measure.

Intuitively, the Fréchet distance d_F is often described through the dog on a leash analogy. Imagine a person walking along a curve while holding the leash of a dog walking along another curve. Both the person and the dog can traverse their respective curves at varying speeds, but can never move backwards. Then, the Fréchet distance is given as the length of

²The function d is said to be a *semimetric* if it fulfils items (i), (ii), and (iii) of Definition 1.1.1, but not item (iv), the triangle inequality.

the shortest possible such leash that still allows both the person and the dog to complete the walk along the curves from the start to the end. To return to the above example, the Fréchet distance between the two curves $d_F(f, g) = 2$ as, due to the opposite phases, the start points of the curves are 1 and -1 respectively.

The use of the Fréchet distance instead of the Hausdorff distance in shape and pattern matching applications for (polygonal) curves was proposed by Alt and Godau [1992], who also propose an algorithm for the computation of the exact Fréchet distance. A close discrete approximation of the Fréchet distance, called the coupling distance, was developed by Eiter and Mannila [1994], who showed that it can be computed efficiently using a very simple algorithm. A more computationally efficient algorithm for the exact Fréchet distance was developed by Alt and Godau [1995], and algorithms for the computation of the exact Fréchet distance under transformations, such as translation or translation with rotation, were developed by Alt et al. [2001] and Wenk [2003] respectively. Among other curve shape matching applications, the Fréchet distance has been successfully applied in vehicle tracking and GPS systems (Alt et al. [2003], Brakatsoulas et al. [2005]) and in molecular biology in protein structure alignment (Aronov et al. [2006], Jiang et al. [2008]).

In the context of continuous functions \mathcal{F} , the Fréchet distance between two functions $f, g \in \mathcal{F}$ is defined as the Fréchet distance between their curves C_f and C_g on $[0, 1] \times \mathbb{R}^m$. For the classical (supremum) Fréchet distance, there are two definitions; reparametrization with respect to one or both of the two curves. These are henceforth called the one-sided Fréchet distance and the two-sided Fréchet distance.

Definition 3.2.1 (One-sided Fréchet distance). Let Φ be the space of all continuous monotonically increasing functions $\phi : [0, 1] \mapsto [0, 1]$ with $\phi(0) = 0$ and $\phi(1) = 1$. Then the one-sided Fréchet distance d_{oF} between the functions $f, g \in \mathcal{F}$ is defined as

$$d_{oF}(f, g) := \inf_{\phi \in \Phi} \sup_{t \in [0, 1]} d_C(C_f(t), C_g(\phi(t))).$$

Definition 3.2.2 (Two-sided Fréchet distance). Let Φ be the space of all continuous monotonically increasing functions $\phi : [0, 1] \mapsto [0, 1]$ with $\phi(0) = 0$ and $\phi(1) = 1$. Then the two-sided Fréchet distance d_{tF} between the functions $f, g \in \mathcal{F}$ is defined as

$$d_{tF}(f, g) := \inf_{\alpha, \beta \in \Phi} \sup_{t \in [0, 1]} d_C(C_f(\alpha(t)), C_g(\beta(t))).$$

A weaker version of the Fréchet distance, d_{wF} , can be attained from the two-sided definition by considering the infima over all continuous *surjective* functions instead. As the Fréchet distance (in both of the above definitions) defines a proper metric for general curves, it thus also defines a proper metric for curves arising from functions.

Note that in the classical supremum sense, the two definitions are equivalent; For all $\theta \in \Phi$ there exists some $s, t \in [0, 1]$ such that

$$d_C(C_f(\alpha(s)), C_g(\beta(s))) = d_C(C_f(\alpha \circ \theta(t)), C_g(\beta \circ \theta(t))).$$

By choosing $\theta = \alpha^{-1}$ and denoting $\phi = \beta \circ \alpha^{-1}$, we get

$$d_C(C_f(t), C_g(\beta \circ \alpha^{-1}(t))) = d_C(C_f(t), C_g(\phi(t))).$$

However, as we will later notice, these definitions are no longer equivalent when embedded in to the integral context.

In the multivariate case, $f : [0, 1]^k \mapsto \mathbb{R}^n$ where $k \leq n$, the Fréchet distance between surfaces is usually attained from the one-sided definition by considering the infima over all orientation preserving homeomorphisms, but definitions with general homeomorphisms have also been considered in the literature. (See for example Fréchet [1924] and Radó [1948] for more details.) While polynomial-time algorithms are known for computation of the Fréchet distance between curves, the underlying decision problem for surfaces is known to be NP-hard (Godau [1998]).

Remark 3.2.1. Naively, one may also define the Fréchet distance for functions $f, g \in \mathcal{F}$ directly between the functions instead of their curves;

$$d_{nF}(f, g) := \inf_{\phi \in \Phi} \sup_{t \in [0, 1]} d_V(f(t), g(\phi(t))).$$

However, what is obtained in this case is only a pseudometric for the functions in \mathcal{F} , that becomes a proper metric when one considers the reparametrizations of f with respect to all $\phi \in \Phi$ as equivalence classes.

3.2.1 Integrated Fréchet distance

In the literature, there have been some attempts at developing an integrated version of the Fréchet distance. For instance, Buchin [2007] considered numerous different approaches to a summed and integrated Fréchet distance, and studied their properties in great detail. Efrat et al. [2007] developed a notion of dynamic time warping, adapted to the continuous context of curves, that bears close resemblance to the notion of integrated Fréchet distance. Buchin et al. [2009] considered a measure for the partial similarity of curves, based on matching pieces of the curves within a fixed Fréchet distance threshold.

Surprisingly, however, formulating a rigorous definition for an integrated Fréchet distance has turned out to not be as straightforward as in the Hausdorff case, and many of the intuitive approaches fail to satisfy some of the key properties of a metric. Specifically, the triangle inequality commonly fails to hold for many of the approaches considered previously in the literature (see Buchin [2007]). In the following, we discuss some

intuitive approaches and provide insight into the difficulties in developing a rigorous definition. For further discussion, we refer the reader to the excellent analysis provided by Buchin [2007] on the integrated Fréchet distance.

We start with the naive definitions, derived directly from Definitions 3.2.1 and 3.2.2 of the one- and two-sided Fréchet distances by substituting the supremum with an integral:

Definition 3.2.3 (Naive one-sided integrated Fréchet distance). Let Φ be the space of all continuous monotonically increasing functions $\phi : [0, 1] \mapsto [0, 1]$ with $\phi(0) = 0$ and $\phi(1) = 1$. The naive one-sided integrated Fréchet distance d_{noIF} between the functions $f, g \in \mathcal{F}$ is defined as

$$d_{noIF}(f, g) := \inf_{\phi \in \Phi} \int_0^1 d_C(C_f(t), C_g(\phi(t))) dt.$$

Definition 3.2.4 (Naive two-sided integrated Fréchet distance). Let Φ be the space of all continuous monotonically increasing functions $\phi : [0, 1] \mapsto [0, 1]$ with $\phi(0) = 0$ and $\phi(1) = 1$. The naive two-sided integrated Fréchet distance d_{ntIF} between the functions $f, g \in \mathcal{F}$ is defined as

$$d_{ntIF}(f, g) := \inf_{\alpha, \beta \in \Phi} \int_0^1 d_C(C_f(\alpha(t)), C_g(\beta(t))) dt.$$

Remark 3.2.2. While the one- and two-sided Fréchet distances are equivalent in the classical sup-norm sense, as was discussed above in Section 3.2, this equivalence no longer holds true for the naive integrated versions where the supremum is replaced with an integral. With a suitable change of variables in the definitions, it is straightforward to see that the naive two-sided definition corresponds to finding the infimum in a weighted space with respect to a reparametrization *and* a weight function, whereas the one-sided definition is only concerned with the infimum with respect to a reparametrization.

Unfortunately, these naive definitions come with a range of glaring issues and notably, neither of them defines a metric for curves or for functions; The naive two-sided integrated Fréchet distance between the functions f and g degenerates to the minimum distance between their curves, C_f and C_g . Consider the Definition 3.2.4. For differentiable and strictly increasing α , a change of variables $z = \alpha(t)$ gives

$$\begin{aligned} & \int_0^1 d_C(C_f(\alpha(t)), C_g(\beta(t))) dt \\ &= \int_0^1 d_C(C_f(z), C_g(\beta \circ \alpha^{-1}(z))) (\alpha^{-1})'(z) dz \\ &= \int_0^1 d_C(C_f(z), C_g(\phi(z))) \omega(z) dz, \end{aligned}$$

where we rewrite $\phi = \beta \circ \alpha^{-1}$ and $\omega = (\alpha^{-1})'$. Here ϕ is continuous and monotonically increasing with $\phi(0) = 0$ and $\phi(1) = 1$, and ω is a weight function on $[0, 1]$, i.e. $\omega \geq 0$ and $\int_0^1 \omega(z) dz = 1$. Hence, by a simple approximation argument, we deduce that

$$\begin{aligned} d_{ntIF}(f, g) &= \inf_{\alpha, \beta \in \Phi} \int_0^1 d_C(C_f(\alpha(t)), C_g(\beta(t))) dt \\ &= \inf_{\omega \in \Omega, \phi \in \Phi} \int_0^1 d_C(C_f(z), C_g(\phi(z))) \omega(z) dz, \end{aligned}$$

where Ω denotes the set of weight functions on $[0, 1]$. Consider now a sequence $\omega_n \rightarrow \delta_{t_0}$ as $n \rightarrow \infty$, that approximates the Dirac-delta function δ_{t_0} at t_0 . Then, as $n \rightarrow \infty$ we have

$$\int_0^1 d_C(C_f(z), C_g(\phi(z))) \omega_n(z) dz \rightarrow d_C(C_f(t_0), C_g(\phi(t_0))).$$

Thus, d_{ntIF} becomes the smallest point-wise distance between the two curves:

$$\begin{aligned} d_{ntIF}(f, g) &= \inf_{\omega \in \Omega, \phi \in \Phi} \int_0^1 d_C(C_f(z), C_g(\phi(z))) \omega(z) dz \\ &\leq \inf_{t_0 \in [0, 1], \phi \in \Phi} d_C(C_f(t_0), C_g(\phi(t_0))). \end{aligned}$$

In particular, if the curves of the functions run through the same point anywhere on $[0, 1] \times \mathbb{R}^m$, the distance d_{ntIF} between them is zero.

From this it is also straightforward to see that d_{ntIF} does not satisfy the triangle inequality; Choose f, g and h such that C_f and C_g never intersect, but C_h intersects with the curves of the other two functions at different points. Then $d_{ntIF}(f, g) > 0$, but $d_{ntIF}(f, h) = d_{ntIF}(h, g) = 0$.

The naive one-sided Fréchet distance d_{noIF} , on the other hand, is not symmetric and does not satisfy the triangle inequality either. Indeed, by very similar arguments as for the naive two-sided version, we obtain

$$\begin{aligned} d_{noIF}(f, g) &= \inf_{\phi \in \Phi} \int_0^1 d_C(C_f(t), C_g(\phi(t))) dt \\ &= \inf_{\phi \in \Phi} \int_0^1 d_C(C_f(\phi^{-1}(z)), C_g(z)) (\phi^{-1})'(z) dz \\ &= \inf_{\theta \in \Phi} \int_0^1 d_C(C_f(\theta(z)), C_g(z)) \theta'(z) dz \end{aligned}$$

which clearly, in general, differs from

$$d_{noIF}(g, f) = \inf_{\phi \in \Phi} \int_0^1 d_C(C_f(\phi(t)), C_g(t)) dt.$$

We also see that the triangle inequality does not hold for d_{noIF} , by again choosing a sequence θ_n such that $\theta'_n \rightarrow \delta_{t_0}$ as $n \rightarrow \infty$, the Dirac-delta function δ_{t_0} at t_0 . By denoting $\theta_\infty = \lim_{n \rightarrow \infty} \theta_n$, the step function corresponding to δ_{t_0} , we have

$$\int_0^1 d_C(C_f(\theta_n(z)), C_g(z))\theta'_n(z)dz \rightarrow d_C(C_f(\theta_\infty(t_0)), C_g(t_0)).$$

Here, θ_∞ can be approximated asymmetrically such that $\theta_\infty(t_0) = t_0$ for any point t_0 . Thus, if the curves of the functions intersect at any point t_0 , the resulting d_{noIF} is zero. Choosing functions f , g , and h as previously with the naive two-sided definition provides a counter-example for triangle inequality in the naive one-sided case as well.

Buchin [2007] discusses two intuitive approaches to developing a general integrated Fréchet distance. Namely, as the integral over a monotone path either in the free space or in the image space. Recall that the path integral of a curve $F : \mathbb{R}^2 \mapsto \mathbb{R}$ over a piecewise continuously differentiable path $\gamma : [0, 1] \mapsto \mathbb{R}^2$ is

$$\int_\gamma F(z)dz = \int_0^1 F(\gamma(t))|\gamma'(t)|dt,$$

where γ' denotes the derivative of γ .

Consider the free space diagram; For $\epsilon > 0$ the free space diagram of two curves C_f and C_g is defined as

$$F_\epsilon(C_f, C_g) := \{(s, t) \in [0, 1]^2 : d_C(C_f(s), C_g(t)) \leq \epsilon\}.$$

The Fréchet distance between two curves is at most ϵ if there exists a monotone path from $(0, 0)$ to $(1, 1)$, in the free space diagram of the curves. Consider a path $\gamma(t) = (t, \phi(t))$ in the free space, with $|\gamma'(t)| = \sqrt{1 + \phi'^2(t)}$. Let $\|\cdot\|_C$ denote the underlying norm on $[0, 1] \times \mathbb{R}^m$. Then, the integrated Fréchet distance can be defined as the infimum of integrals over such paths as follows:

Definition 3.2.5 (Free space integrated Fréchet distance). Let Φ be the space of all continuous monotonically increasing functions $\phi : [0, 1] \mapsto [0, 1]$ with $\phi(0) = 0$ and $\phi(1) = 1$. Then the free space integrated Fréchet distance d_{fIF} between the functions $f, g \in \mathcal{F}$ is defined as

$$d_{fIF}(f, g) := \inf_{\phi \in \Phi} \int_0^1 \|C_f(t) - C_g(\phi(t))\|_C \sqrt{1 + \phi'^2(t)} dt.$$

Alternatively, the integrated Fréchet distance can be defined by taking a path integral over the curves in image space simultaneously. For this, we must assume the curves C_f and C_g to be piecewise continuously differentiable. However, the curves arising from the continuous functions \mathcal{F} satisfy this condition. Denote by C'_f the derivative of the curve C_f . Then, $\|C'_f(\cdot)\|_C$ gives a notion of the incremental arc length traversed along the curve C_f . This leads to the following definition.

Definition 3.2.6 (Image space integral Fréchet). Let Φ be the space of all continuous monotonically increasing functions $\phi : [0, 1] \mapsto [0, 1]$ with $\phi(0) = 0$ and $\phi(1) = 1$. Then the image space integrated Fréchet distance d_{iIF} between the functions $f, g \in \mathcal{F}$ is defined as

$$d_{iIF}(f, g) := \inf_{\phi \in \Phi} \int_0^1 \|C_f(t) - C_g(\phi(t))\|_C (\|C'_f(t)\|_C + \|(C_g \circ \phi)'(t)\|_C) dt.$$

Remark 3.2.3. While Buchin [2007] introduces the above two ideas in the context of normed spaces, they generalize to metric spaces as well. In both Definitions 3.2.5 and 3.2.6, replace the norm $\|C_f(t) - C_g(\phi(t))\|_C$ with the metric $d_C(C_f(t), C_g(\phi(t)))$, and in Definition 3.2.6 replace $\|C'_f\|_C$ by some d_C -based function $L : [0, 1] \times \mathbb{R}^m \mapsto [0, \infty)$ that gives a notion of the incremental arc length traversed along the curve.

Note that in order to be formally correct in the above definitions, one should take the infimum over piecewise continuously differentiable functions $\phi \in \Phi$. This minor issue however does not play any fundamental role. While the above definitions give *summed* Fréchet distances, Buchin [2007] also develops several slight variations that yield *averaged* Fréchet distances. Although these approaches seem intuitive, the detailed analysis provided in Buchin [2007] shows that they do not satisfy the triangle inequality for general curves in \mathbb{R}^m . While the simple counterexamples provided by Buchin [2007] disproving the triangle inequality for Definitions 3.2.5 and 3.2.6 do not belong in the class of curves arising from functions, they lead to suspect that such functions for which the triangle inequality doesn't hold can also be found.

4. Summaries of the articles

The main scientific contribution of this thesis lies within the three publications that follow. Summaries of the key contributions are given below.

I. Pareto depth for functional data

Due to the richness of functional data, it is incredibly difficult to provide a notion of depth that simultaneously captures the features of all of the important modes of variation (in location, shape, and structure) of a functional distribution. Therefore, we take a data-driven approach to functional depth that allows for great flexibility in the choice of the incorporated features.

In the approach, the functional observations $x \in \mathcal{H}$ are mapped to J -variate vectors through the use of several mappings $s_j : \mathcal{H} \rightarrow \mathbb{R}$, $j = 1, \dots, J$, called the *statistics of interest* (SOI), that quantify some inherent features of the distribution $P \in \mathcal{P}(\mathcal{H})$. Then, the functional observations are assigned depth values by computing a new multivariate depth, the *Pareto Depth*, on the vectors of SOI. This approach allows for the analyst to leverage their knowledge of the context of the data and application of the method in choosing the set of SOI. In particular, the proposed method makes no assumptions on the types of mappings used in capturing the essential modes of variation in the data. Many other dimension-reduction type methodologies in functional data analysis, based on replacing the random function x with a (random) vector, often assume that the interesting characteristics of P can be captured through linear projections of x on to suitable elements of \mathcal{H} . This, however, need not be the case, and the methodology adopted in the article allows for greater creativity in choosing the mappings appropriate to the context. The proposed multivariate depth method, the Pareto depth, is then purposefully built on componentwise comparison of the vectors of SOI, as the geometry of the underlying multivariate space is of no interest due to the potentially wildly differing types of mappings (the individual SOI) the margins of the space describe.

We provide results on the consistency of the proposed multivariate Pareto depth method and explore the properties of the corresponding functional depth approach. We discuss the choice of the statistics of interest and study the application of the functional depth approach in practice. We illustrate the advantages of such a flexible choice of features in several simulated and real data examples. In particular, we demonstrate the excellent performance of the proposed approach in applications such as classification, where the performance of the applied method is directly tied to its ability to capture the appropriate factors that discriminate between the distributions of the compared classes.

II. Flexible integrated functional depths

We introduce a new class of functional depths, called the J th order k th moment integrated depths, and study their properties. The introduced approach expands upon the ideas considered by Nagy et al. [2017].

For $x \in \mathcal{H}(\mathcal{V})$ and $P \in \mathcal{P}(\mathcal{H})$, consider the distribution of multivariate depths of the joint vector of J time margins $t_1, \dots, t_J \in \mathcal{V}$, with respect to the corresponding joint distribution of the margins. That is, consider the distribution of $D((x(t_1), \dots, x(t_J))^T, P_{(X(t_1), \dots, X(t_J))^T})$ over $t_1, \dots, t_J \in \mathcal{V}$, where $(x(t_1), \dots, x(t_J))^T \in \mathbb{R}^J$, $P_{(X(t_1), \dots, X(t_J))^T} \in \mathcal{P}(\mathbb{R}^J)$ and $X \sim P$. The J th order k th moment integrated depth is then defined through the k th moment of this multivariate depth distribution. Due to theoretical reasons, in the paper, we restrict to considering the multivariate halfspace depth. However, in principle, any multivariate depth could be utilized, and the construction would provide a corresponding integrated functional version of the depth.

We provide several interesting asymptotic properties of the proposed class of depths. We show that the proposed depths are uniformly consistent, and derive their limiting distributions under mild regularity conditions. We study the properties of the proposed depth approach in practice in several real data examples. In particular, we illustrate the versatility of the approach in automatically capturing the important features of the functional distributions in an application to supervised classification.

III. Integrated shape-sensitive functional metrics

Sensitivity of the methodology to variations in the shape of functions has been a commonly discussed problem in many fields of research. Popular approaches to metricising the shape similarity between functions include considering shape-sensitive metrics, such as Hausdorff or Fréchet distances, between the graphs or the curves of the functions. However, as

both of these distances are based on a global supremum of a pointwise distance between the graphs, they do not consider the general shape similarity of the functions. Thus, in the literature, some attention has been devoted to the search of averaged (i.e. integrated) versions of these metrics, that better reflect the overall local likeness of the graphs. Although some integrated versions of these metrics have been proposed in the literature, none of the previously considered approaches provide a metric suitable to the context of functional data, without severe drawbacks or limitations (as demonstrated in Section 3).

In the paper, we introduce a new metric construction, the integrated ball (pseudo)metric $d^{\epsilon,p}$, that provides an integrated version of any (pseudo)metric d . In the construction, the distance d is computed locally, between the pieces of the graphs when restricted to neighbourhoods of size ϵ . The local neighbourhoods are then integrated over the domain within the L_p norm, providing an overall measure of the local likeness (based on metric d) between the graphs. In particular, we apply the integrated ball construction to the Hausdorff and Fréchet distances to provide their integrated ball versions. To the best of our knowledge, the integrated ball Hausdorff and Fréchet distances appear to be the first integrated versions of these distances, that provide continuous metrics suitable to the context of functional data.

We show that under some very mild and natural assumptions, for any metric d , the integrated ball construction $d^{\epsilon,p}$ produces a metric, and has some very important continuity and consistency properties. Furthermore, we show that the Hausdorff and Fréchet distances (when suitably adjusted to the context of local restrictions of graphs) fulfill all of our assumptions, thus ensuring that their integrated ball counterparts enjoy all of the desirable properties that are shown to hold for the integrated ball metrics. Finally, we study the properties of the integrated ball Hausdorff and Fréchet distances in practice, in several simulated outlier detection problems based on a difference in shape. In the simulation study, we demonstrate the benefits of assessing the overall local likeness of the curves over a range of locality levels ϵ .

References

- C. Agostinelli and M. Romanazzi. Local depth. *Journal of Statistical Planning and Inference*, 141(2):817–830, 2011.
- H. Alt and M. Godau. Measuring the resemblance of polygonal curves. In *Proceedings of the Eighth Annual Symposium on Computational Geometry*, pages 102–109, 1992.
- H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 5(01n02):75–91, 1995.
- H. Alt, C. Knauer, and C. Wenk. Matching polygonal curves with respect to the Fréchet distance. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 63–74. Springer, Berlin, Heidelberg, 2001.
- H. Alt, A. Efrat, G. Rote, and C. Wenk. Matching planar maps. *Journal of Algorithms*, 49(2):262–283, 2003.
- B. Aronov, S. Har-Peled, C. Knauer, Y. Wang, and C. Wenk. Fréchet distance for curves, revisited. In *European Symposium on Algorithms*, pages 52–63. Springer, Berlin, Heidelberg, 2006.
- B. Aulbach, M. Rasmussen, and S. Siegmund. Approximation of attractors of nonautonomous dynamical systems. *Discrete & Continuous Dynamical Systems-B*, 5(2):215–238, 2005.
- A. J. Baddeley. Errors in binary images and an L_p version of the Hausdorff metric. *Nieuw Archief voor Wiskunde*, 10(4):157–183, 1992.
- V. Barnett. The ordering of multivariate data. *Journal of the Royal Statistical Society: Series A (General)*, 139(3):318–344, 1976.
- J. M. Bogoya, A. Vargas, O. Cuate, and O. Schütze. A (p, q) -averaged Hausdorff distance for arbitrary measurable sets. *Mathematical and Computational Applications*, 23(3):51, 2018.
- D. Bosq. *Linear processes in function spaces: theory and applications*. Springer Science & Business Media, 2000.
- S. Brakatsoulas, D. Pfoser, R. Salas, and C. Wenk. On map-matching vehicle tracking data. In *Proceedings of the 31st International Conference on Very Large Data Bases*, pages 853–864, 2005.
- K. Buchin, M. Buchin, and Y. Wang. Exact algorithms for partial curve matching via the Fréchet distance. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 645–654. Society for Industrial and Applied Mathematics, 2009.

- M. E. Buchin. *On the computability of the Fréchet distance between triangulated surfaces*. PhD thesis, Freie Universität Berlin, Germany, 2007.
- W. B. Capra and H. G. Müller. An accelerated-time model for response curves. *Journal of the American Statistical Association*, 92(437):72–83, 1997.
- A. Chakraborty and P. Chaudhuri. On data depth in infinite dimensional spaces. *Annals of the Institute of Statistical Mathematics*, 66(2):303–324, 2014a.
- A. Chakraborty and P. Chaudhuri. The spatial distribution in infinite dimensional spaces and related quantiles and depths. *The Annals of Statistics*, 42(3):1203–1231, 2014b.
- P. Chaudhuri. On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872, 1996.
- Y. Chen, X. Dang, H. Peng, and H. L. Bart. Outlier detection with the kernelized spatial depth function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):288–305, 2009.
- G. Claeskens, M. Hubert, L. Slaets, and K. Vakili. Multivariate functional half-space depth. *Journal of the American Statistical Association*, 109(505):411–423, 2014.
- J. A. Cuesta-Albertos and A. Nieto-Reyes. The random Tukey depth. *Computational Statistics & Data Analysis*, 52(11):4979–4988, 2008.
- A. Cuevas and R. Fraiman. On depth measures and dual statistics. a methodology for dealing with general data. *Journal of Multivariate Analysis*, 100(4):753–766, 2009.
- A. Cuevas, M. Febrero, and R. Fraiman. Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 23(3):481–496, 2007.
- I. Daubechies. *Ten lectures on wavelets*. Society for industrial and applied mathematics, 1992.
- C. de Boor. *A practical guide to splines. Revised edition*. New York: Springer-Verlag, 2001.
- F. D. A. De Carvalho, R. M. de Souza, M. Chavent, and Y. Lechevallier. Adaptive Hausdorff distances and dynamic clustering of symbolic interval data. *Pattern Recognition Letters*, 27(3):167–179, 2006.
- M. Dellnitz and A. Hohmann. A subdivision algorithm for the computation of unstable manifolds and global attractors. *Numerische Mathematik*, 75(3):293–317, 1997.
- M. Dellnitz, O. Schütze, and T. Hestermeyer. Covering Pareto sets by multilevel subdivision techniques. *Journal of Optimization Theory and Applications*, 124(1):113–136, 2005.
- D. L. Donoho and M. Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803–1827, 1992.
- S. Dutta and A. K. Ghosh. On robust classification using projection depth. *Annals of the Institute of Statistical Mathematics*, 64(3):657–676, 2012.
- A. Efrat, Q. Fan, and S. Venkatasubramanian. Curve matching, time warping, and light fields: New algorithms for computing similarity between curves. *Journal of Mathematical Imaging and Vision*, 27(3):203–216, 2007.

- T. Eiter and H. Mannila. Computing discrete Fréchet distance. Technical Report CD-TR 94/64, Christian Doppler Laboratory for Expert Systems, TU Vienna, Austria, 1994.
- I. Epifanio. Shape descriptors for classification of functional data. *Technometrics*, 50(3):284–294, 2008.
- K. Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.
- J. Fan and I. Gijbels. *Local polynomial modelling and its applications*. Routledge, 1996.
- F. Ferraty and P. Vieu. *Nonparametric functional data analysis: theory and practice*. New York: Springer, 2006.
- R. Fraiman and J. Meloche. Multivariate L-estimation. *Test*, 8:255–317, 1999.
- R. Fraiman and G. Muniz. Trimmed means for functional data. *Test*, 10(2): 419–440, 2001.
- M. Fréchet. Sur la distance de deux surface. *Annales de la Société Polonaise de Mathématique*, 3:4–19, 1924.
- A. K. Ghosh and P. Chaudhuri. On maximum depth and related classifiers. *Scandinavian Journal of Statistics*, 32(2):327–350, 2005a.
- A. K. Ghosh and P. Chaudhuri. On data depth and distribution-free discriminant analysis using separating surfaces. *Bernoulli*, 11(1):1–27, 2005b.
- I. Gijbels and S. Nagy. On a general definition of depth for functional data. *Statistical Science*, 32(4):630–639, 2017.
- M. Godau. *On the complexity of measuring the similarity between geometric objects in higher dimensions*. PhD thesis, Freie Universität Berlin, Germany, 1998.
- A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910.
- P. Hall, H. G. Müller, and J. L. Wang. Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 34(3): 1493–1517, 2006.
- M. Hallin, D. Paindaveine, and M. Šíman. Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *The Annals of Statistics*, 38(2):635–669, 2010.
- T. Harris, J. D. Tucker, B. Li, and L. Shand. Elastic depths for detecting shape anomalies in functional data. *Technometrics*, 63(4):466–476, 2021.
- T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. New York: springer, 2009.
- S. Helander, G. Van Bever, S. Rantala, and P. Ilmonen. Pareto depth for functional data. *Statistics*, 54(1):182–204, 2020.
- S. Helander, P. Laketa, P. Ilmonen, S. Nagy, G. Van Bever, and L. Viitasaari. Integrated shape-sensitive functional metrics. *Journal of Multivariate Analysis*, 189:104880, 2021.
- D. Hlubinka, L. Kotík, and O. Vencálek. Weighted halfspace depth. *Kybernetika*, 46(1):125–148, 2010.

- L. Horváth and P. Kokoszka. *Inference for functional data with applications*. Springer Science & Business Media, 2012.
- T. Hsing and R. Eubank. *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons, 2015.
- D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.
- J. Jacques and C. Preda. Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8(3):231–255, 2014.
- M. Jiang, Y. Xu, and B. Zhu. Protein structure–structure alignment with discrete Fréchet distance. *Journal of Bioinformatics and Computational Biology*, 6(1): 51–64, 2008.
- N. M. Jouzdani and V. M. Panaretos. Functional data analysis with rough sampled paths? *arXiv preprint*, arXiv:2105.12035, 2021.
- G. Koshevoy and K. Mosler. Zonoid trimming for multivariate distributions. *The Annals of Statistics*, 25(5):1998–2017, 1997.
- J. Li, J. A. Cuesta-Albertos, and R. Liu. DD-classifier: Nonparametric classification procedure based on DD-plot. *Journal of the American statistical association*, 107(498):737–753, 2012.
- Y. Li and T. Hsing. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38(6):3321–3351, 2010.
- R. Liu. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414, 1990.
- R. Liu and K. Singh. A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88(421):252–260, 1993.
- R. Liu, J. M. Parelius, and K. Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion). *The Annals of Statistics*, 27:783–858, 1999.
- S. López-Pintado and J. Romo. On the concept of depth for functional data. *Journal of the American statistical Association*, 104(486):718–734, 2009.
- S. López-Pintado and J. Romo. A half-region depth for functional data. *Computational Statistics & Data Analysis*, 55(4):1679–1695, 2011.
- P. C. Mahalanobis. On the generalized distance in statistics. *National Institute of Science of India*, 12:49–55, 1936.
- J. S. Marron and A. B. Tsybakov. Visual error criteria for qualitative smoothing. *Journal of the American Statistical Association*, 90(430):499–507, 1995.
- J. S. Marron, J. O. Ramsay, L. M. Sangalli, and A. Srivastava. Functional data analysis of amplitude and phase variation. *Statistical Science*, pages 468–484, 2015.
- J. C. Massé and R. Theodorescu. Halfplane trimming for bivariate distributions. *Journal of Multivariate Analysis*, 48(2):188–202, 1994.
- K. Mosler. Depth statistics. In *Robustness and complex data structures*, pages 17–34. Springer, Berlin, Heidelberg., 2013.

- S. Nagy and J. Dvořák. Illumination depth. *Journal of Computational and Graphical Statistics*, 30(1):78–90, 2021.
- S. Nagy, I. Gijbels, M. Omelka, and D. Hlubinka. Integrated depth for functional data: statistical properties and consistency. *ESAIM: Probability and Statistics*, 20:95–130, 2016.
- S. Nagy, I. Gijbels, and D. Hlubinka. Depth-based recognition of shape outlying functions. *Journal of Computational and Graphical Statistics*, 26(4):883–893, 2017.
- S. Nagy, C. Schütt, and E. M. Werner. Halfspace depth and floating body. *Statistics Surveys*, 13:52–118, 2019.
- S. Nagy, S. Helander, G. Van Bever, L. Viitasaari, and P. Ilmonen. Flexible integrated functional depths. *Bernoulli*, 27(1):673–701, 2021.
- A. Nieto-Reyes. On the properties of functional depth. In *Recent Advances in Functional Data Analysis and Related Topics; Selected papers from the 2nd International Workshop on Functional and Operational Statistics (IWFOS 2011) (F. Ferraty, ed.)*, page 239–244. Physica-Verlag / Springer, Heidelberg, 2011.
- A. Nieto-Reyes and H. Battey. A topologically valid definition of depth for functional data. *Statistical Science*, 31(1):61–79, 2016.
- D. Nolan. Asymptotics for multivariate trimming. *Stochastic processes and their applications*, 42(1):157–169, 1992.
- H. Oja. Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1(6):327–332, 1983.
- D. Paindaveine and G. Van Bever. From depth to local depth: a focus on centrality. *Journal of the American Statistical Association*, 108(503):1105–1119, 2013.
- D. Paindaveine and G. Van Bever. Nonparametrically consistent depth-based classifiers. *Bernoulli*, 21(1):62–82, 2015.
- C. Preda. Regression models for functional data by reproducing kernel Hilbert spaces methods. *Journal of statistical planning and inference*, 137(3):829–840, 2007.
- T. Radó. *Length and Area*, volume 30. American Mathematical Soc., 1948.
- J. O. Ramsay and X. Li. Curve registration. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(2):351–363, 1998.
- J. O. Ramsay and B. Silverman. *Functional data analysis, 2nd Ed.* Springer-Verlag, 2005.
- J. O. Ramsay, R. D. Bock, and T. Gasser. Comparison of height acceleration curves in the Fels, Zurich, and Berkeley growth data. *Annals of Human Biology*, 22(5):413–426, 1995.
- J. A. Rice and B. W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):233–243, 1991.
- P. J. Rousseeuw and M. Hubert. Regression depth. *Journal of the American Statistical Association*, 94(446):388–402, 1999.
- P. J. Rousseeuw and I. Ruts. Algorithm as 307: Bivariate location depth. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(4):516–526, 1996.

- P. J. Rousseeuw and I. Ruts. The depth function of a population distribution. *Metrika*, 49(3):213–244, 1999.
- P. J. Rousseeuw and A. Struyf. Computing location depth and regression depth in higher dimensions. *Statistics and Computing*, 8(3):193–203, 1998.
- W. J. Rucklidge. Efficiently locating objects using the Hausdorff distance. *International Journal of Computer Vision*, 24(3):251–270, 1997.
- I. Ruts and P. J. Rousseeuw. Computing depth contours of bivariate point clouds. *Computational statistics & data analysis*, 23(1):153–168, 1996.
- O. Schütze. *Set oriented methods for global optimization*. PhD thesis, University of Paderborn, 2004.
- O. Schütze, M. Laumanns, C. A. C. Coello, M. Dellnitz, and E. G. Talbi. Convergence of stochastic search algorithms to finite size Pareto set approximations. *Journal of Global Optimization*, 41(4):559–577, 2008.
- O. Schütze, X. Esquivel, A. Lara, and C. A. C. Coello. Using the averaged Hausdorff distance as a performance measure in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 16(4):504–522, 2012.
- R. Serfling. A depth function and a scale curve based on spatial quantiles. In *Statistical data analysis based on the L_1 -norm and related methods*, pages 25–38. Birkhäuser, Basel., 2002.
- R. Serfling. Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *Journal of Nonparametric Statistics*, 27(7):915–936, 2010.
- C. Sguera, P. Galeano, and R. Lillo. Spatial depth-based classification for functional data. *Test*, 23(4):725–750, 2014.
- C. Sguera, P. Galeano, and R. Lillo. Functional outlier detection by a local depth with application to NOx levels. *Stochastic environmental research and risk assessment*, 30(4):1115–1130, 2016.
- K. Singh. A notion of majority depth. Technical report, Rutgers University, Dept. of Statistic, 1991.
- Y. Sun and M. G. Genton. Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334, 2011.
- S. J. Taylor. The exact Hausdorff measure of the sample path for planar Brownian motion. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 60(2), pages 253–258. Cambridge University Press, 1964.
- R. D. Tuddenham. Physical growth of california boys and girls from birth to eighteen years. *University of California publications in child development*, 1: 183–364, 1954.
- J. W. Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. V., 1974)*, volume 2, pages 523–531. Canadian Mathematical Congress, Montreal, Que., 1975.
- A. Vargas and J. M. Bogoya. A generalization of the averaged Hausdorff distance. *Computación y Sistemas*, 22:331–345, 2018.
- R. C. Veltkamp and M. Hagedoorn. State of the art in shape matching. In *Principles of Visual Information Retrieval. Advances in Pattern Recognition.*, pages 87–119. Springer, London, 2001.
- M. P. Wand and M. C. Jones. *Kernel smoothing*. CRC press, 1994.

- C. Wenk. *Shape matching in higher dimensions*. PhD thesis, Freie Universität Berlin, Germany, 2003.
- F. Yao, H. G. Müller, and J. L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.
- A. B. Yeh and K. Singh. Balanced confidence regions based on Tukey’s depth and the bootstrap. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):639–652, 1997.
- X. Yi and O. I. Camps. Line-based recognition using a multidimensional Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):901–916, 1999.
- X. Zhang and J. L. Wang. From sparse to dense functional data and beyond. *The Annals of Statistics*, 44(5):2281–2321, 2016.
- Y. Zuo and R. Serfling. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 2000a.
- Y. Zuo and R. Serfling. Nonparametric notions of multivariate “scatter measure” and “more scattered” based on statistical depth functions. *Journal of Multivariate analysis*, 75(1):62–78, 2000b.



ISBN 978-952-64-0881-1 (printed)
ISBN 978-952-64-0882-8 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Mathematics and Systems Analysis
www.aalto.fi

**BUSINESS +
ECONOMY**

**ART +
DESIGN +
ARCHITECTURE**

**SCIENCE +
TECHNOLOGY**

CROSSOVER

**DOCTORAL
THESES**