A Euclidean Embedding for Computing Persistent Homology with Gaussian Kernels

Jean-Daniel Boissonnat
Université Côte d'Azur, INRIA, Sophia-Antipolis, France

Kunal Dutta
Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, Poland

Abstract

Computing persistent homology of large datasets using Gaussian kernels is useful in the domains of topological data analysis and machine learning as shown by Phillips, Wang and Zheng [SoCG 2015]. However, unlike in the case of persistent homology computation using the Euclidean distance or the k-distance, using Gaussian kernels involves significantly higher overhead, as all distance computations are in terms of the Gaussian kernel distance which is computationally more expensive. Further, most algorithmic implementations (e.g. Gudhi, Ripser, etc.) are based on Euclidean distances, so the question of finding a Euclidean embedding – preferably low-dimensional – that preserves the persistent homology computed with Gaussian kernels, is quite important.

We consider the Gaussian kernel power distance (GKPD) given by Phillips, Wang and Zheng. Given an n-point dataset and a relative error parameter $\varepsilon \in (0, 1]$, we show that the persistent homology of the Čech filtration of the dataset computed using the GKPD can be approximately preserved using $O\left(\varepsilon^{-2} \log n\right)$ dimensions, under a *high stable rank* condition. Our results also extend to the Delaunay filtration and the (simpler) case of the weighted Rips filtrations constructed using the GKPD.

Compared to the Euclidean embedding for the Gaussian kernel function in $\sim n$ dimensions, which uses the Cholesky decomposition of the matrix of the kernel function applied to all pairs of data points, our embedding may also be viewed as dimensionality reduction – reducing the dimensionality from n to $\sim \log n$ dimensions.

Our proof utilizes the embedding of Chen and Phillips [ALT 2017], based on the *Random Fourier Functions* of Rahimi and Recht [NeurIPS 2007], together with two novel ingredients. The first one is a new decomposition of the squared radii of Čech simplices computed using the GKPD, in terms of the pairwise GKPDs between the vertices, which we state and prove. The second is a new concentration inequality for sums of cosine functions of Gaussian random vectors, which we call *Gaussian cosine chaoses*. We believe these are of independent interest and will find other applications in future.

2012 ACM Subject Classification Theory of computation → Computational geometry; Theory of computation → Random projections and metric embeddings; Theory of computation → Gaussian processes

Keywords and phrases Persistent homology, Gaussian kernels, Random Fourier Features, Euclidean embedding

Digital Object Identifier 10.4230/LIPIcs.ESA.2024.29

Funding *Kunal Dutta*: Supported by the Polish NCN SONATA Grant no. 2019/35/D/ST6/04525.

Acknowledgements The authors thank the referees for numerous helpful suggestions and remarks, which helped improve the presentation and flow of the paper. Special thanks go also to the anonymous referee of a previous version, for pointing out a flaw, correcting which led to a significant improvement in the present results.

1 Introduction

Persistent homology (PH) is one of the main tools to extract information from data in topological data analysis. Given a data set as a point cloud in some ambient space, the idea is to construct a filtration sequence of topological spaces from the point cloud, and extract topological information from this sequence.

Two main issues are to be faced. First, the data points often live in a very high dimensional space and computing PH has exponential or worse dependence on the ambient dimension. It follows that PH rapidly becomes unusable once the dimension grows beyond a few dozens – which is indeed the case in many applications, for example in image processing, neuro-biological networks, and data mining (see e.g. Giraud [17]). This phenomenon is often referred to as the curse of dimensionality. A second major difficulty comes from the fact that data is usually corrupted by noise and outliers. Indeed, while the PH (computed using offsets to a distance function) is quite robust to Hausdorff noise, it is not hard to see that the presence of even a single outlier can significantly affect the PH (see e.g. [11]).

1.1 State of the Art

Persistent Homology with Outliers

One approach to circumvent the issue of outliers is to use distance functions that are more robust to outliers, such as the distance-to-a-measure (DTM) and the related k-distance (for finite data sets), proposed recently by Chazal et al. [8]. This approach also has the advantage of compatibility with de-noising techniques such as [9]. Though the DTM-based approach is promising, it tends to come with a significant increase in run-time complexity, and several approximation-based techniques [18, 8, 2] have been proposed to address this. Another approach to circumvent the issue of outliers, introduced by Phillips et al. [28], involves using *kernels*, which are similarity functions on pairs of points in an ambient space, and widely used in machine learning and data analysis. Computing the PH using the kernel distance has certain advantages compared to the Euclidean or the k-distance, especially for machine learning applications. These include the existence of ε-coresets [20, 26] as well as some properties of the kernel distance function, e.g. correspondence of its sublevel sets to superlevel sets of kernel density estimates, stability with respect to a variable smoothing parameter, and an asymptotic bound for the distance between two measures in terms of the Wasserstein 2-distance. These properties of the Gaussian kernel distance were proved in [28], where an approximate power distance version of the kernel distance – which we call the Gaussian Kernel Power Distance (GKPD) – was used to compute the PH of some datasets and compare with the PH computed using existing distance functions. Further progress in constructing robust persistence diagrams has been recently made in [35] using this approach.

The Kernel Trick and Random Fourier Features

As mentioned earlier, a kernel is a similarity function on pairs of points. Commonly used kernels include the Gaussian kernel (given by $K(x,p) = \exp\{-\|x-p\|^2/2\sigma^2\}$), polynomial kernel, Laplace kernel, Cauchy kernel, etc. Kernels methods are widely used in machine learning, AI and data analysis [19]. Their popularity in these areas owes to the fact that they allow non-linear analysis of the data using linear techniques such as regression. This is accomplished by lifting the input to a higher-dimensional target space where linear methods can be applied. Although in many cases, this target space can have high or even infinite dimension, for a large class of kernels, it is possible to avoid operating directly in this space using the so-called *kernel trick*, which allows inner products in the space to be expressed using only the kernel function in the original space.

Although the kernel trick avoids direct computations in a high-dimensional target space, many kernel based algorithms require operations on the matrix of all pairwise inner products. For instance, a Euclidean embedding approximating the kernel function could be obtained using a Cholesky decomposition of this matrix, but would require $\Omega(n^2)$ operations – untenable for e.g. $n = 10^9$. Further, the embedding would lose the desirable property of data-obliviousness. Making the kernel distance scalable by a kernel distance embedding - preferably data-oblivious, is therefore of significant importance [23, 1].

In this direction, the celebrated result of Rahimi and Recht [30] gives an *approximate* Euclidean embedding which works for a broad class of kernel functions – including Gaussian kernels. Their idea is to use Random Fourier Features (RFFs), which preserve the kernel function up to a $(1 \pm \varepsilon)$-factor. RFFs have since found widespread applicability in machine learning and AI algorithms involving kernel methods [21] – as attested to by more than 4000 citations of [30], as well as the NeurIPS (2017) Test of Time award. Besides learning theory, RFFs for Gaussian kernels have also been used for *metric embeddings* [6], where they are known as *randomized Nash devices*, as they are a randomized version of Nash's embedding in his celebrated proof of the C^1 imbeddability theorem [25].

Euclidean Embedding for Persistent Homology using Kernel Distance

When it comes to computing PH, as discussed previously, using Gaussian kernels – specifically the GKPD – has significant advantages. However, most algorithmic implementations (e.g. Gudhi, Ripser, etc.) work with Euclidean distances and so it becomes important to have a Euclidean embedding for the GKPD, preferably low-dimensional. Observe that directly approximating the matrix of pairwise Gaussian kernel distances would require $O(n^3)$ dimensions. Further, even the RFF map of Rahimi-Recht does not directly work here, as it only approximates the Gaussian kernel *function* $(K(x, y),\ x, y \in \mathbb{R}^D)$, thus giving an additive error for the Gaussian kernel *distance* (given by $\sqrt{2(1 - K(x, y))}$).
To address this issue, recently Chen and Phillips [13] gave a relative approximation of Gaussian kernel distances using RFFs. Further using lower bounds on the well-known Johnson-Lindenstrauss lemma, they also showed that their bound on the number of dimensions used is tight up to a factor of $O(\log 1/\varepsilon)$.

However, another key issue under the GKPD is that the weights associated to the data points are not just a function of the points themselves, but of the pairwise kernel distances of all the points in the data set. This means that for any metric embedding of the data points, the weights of the points *must be recomputed* in the new space, and cannot be simply set to the value of the old weights.

For an approximate embedding, it is therefore necessary to preserve the weight function as well as the pairwise distances between the point. A sufficient condition for this, is to preserve kernel distances between *measures* on the point set. Unfortunately, the Chen-Phillips mapping does not preserve distances between measures, and *a priori* does not preserve the weights of the points. Similarly, other different approaches (e.g. [36, 24, 5, 10]) are not efficient in preserving distances between point sets, or between general measures.

A different embedding, obtained by Phillips and Tai [27] gives a relative approximation with a small additive error for kernel distances between sets of points, though using $O\left(\varepsilon^{-2}\log^2 n\right)$ dimensions (compared to $O\left(\varepsilon^{-2}\log n\right)$ in our case) and a computationally involved implementation. Moreover, even given such a mapping (i.e. preserving kernel distances between point distributions), it is not clear that it can preserve the PH, since this involves preserving intersections of multiple balls under a power distance (see e.g. [4, 28]).

As mentioned in the conclusion of Arya, Boissonnat, Dutta and Lotz [4], it is possible to obtain a constant factor approximation ≥ 4 of the PH, essentially by approximating the kernel distance by the Euclidean distance for small values of the Euclidean distance. However, the question of finding a $(1+\varepsilon)$-factor approximation for PH computation using the GKPD remained open.

1.2 Our Contribution

In this paper, we show that given an n-point dataset P in a space $\mathbb{R}^D$ and any $\varepsilon \in (0, 1]$, it is possible to approximate the PH of P computed using the GKPD, by a $(1+\varepsilon)$-factor together with an additive $o(1)$-factor, using a Euclidean embedding with $O\left(\varepsilon^{-2} \log n\right)$ dimensions – under a *high stable rank* condition.

The formal statement, Theorem 13, is in Section 6. It yields a map allowing us to approximately compute the PH of a set of points in a high dimensional space, using the GKPD, while actually working with Euclidean distances. Thus, we affirmatively answer the question asked in the conclusion of [4].

We also obtain an analogous result for the PH computed over the weighted Rips filtration using the GKPD. Our results are in a sense analogous to the dimensionality reduction results of Sheehy [31] and Lotz [22] for PH using Euclidean distances, and Arya et al. [4] for the Euclidean k-distance.

From the lower bound result of Chen and Phillips [13][Section 6] it can also be seen that the number of dimensions used in their Gaussian kernel distance approximation is tight up to a factor of $O(\log 1/\varepsilon)$. This implies our target dimension is also tight up to an $O(\log 1/\varepsilon)$ factor.

Our results require a stable rank condition on certain matrices formed using the position vectors of the points, where the stable rank of a matrix is the squared ratio of its Frobenius norm to its operator norm [34]. Hence, our high stable rank condition may be interpreted as requiring the pairwise difference vectors of the data points to be well-spread in a subspace of at least $\omega(1)$ dimensions. For high-dimensional datasets – typical in machine learning applications – where the ambient dimension can be of the order of the size of the dataset, it is natural to assume that the intrinsic dimension is not arbitrarily small and is $\omega(1)$. For example, the dimensionality reduction using random projections reduces the data to $\Omega\left(\varepsilon^{-2} \log n\right)$ dimensions, which is tight. This indicates that for some datasets, the intrinsic dimension can be as high as $\log n$. In contrast, we assume a much weaker lower bound of $\omega(1)$.

For more information on the stable rank as well as further applications, we refer the interested reader to Vershynin [34][Chapter 7].

New Tools

Our result is based on two main new tools – one geometric and the other probabilistic. On the geometric side, we give a new geometric decomposition in Section 5 (Simplex Distortion Lemma 12) showing that the distortion of the squared radius of a minimum enclosing ball of a set of weighted points, computed using the GKPD, can be expressed as a linear combination of the distortion of the pairwise power distances between the points. We shall show that although the squared GKPD is non-linear, when lifted to a certain Hilbert space, it has several nice properties, which we then use to prove the lemma.

However, the above decomposition is still not sufficient to show that the weights of the points will also be approximately preserved. This is because, as we shall see, each weight is the *difference* of two mean squared kernel distances. Thus if each pairwise kernel distance is approximately preserved this only preserves their sum and not the difference.

To handle this, on the probabilistic side we prove a new concentration inequality (Section 4, Theorem 8) for a class of *trigonometric functions of Gaussian random vectors*. Under a condition on the matrix having its columns as the vectors $p \in P$, together with a bound on the Euclidean norm of the vectors, we show that the *average* squared kernel distance between pairs of input data points is preserved with a $(1 + o(1))$-factor distortion, and therefore the weight of each point is actually $(1 \pm \varepsilon)$-preserved alongwith a $o(1)$-additive factor.

Our new concentration inequality holds for sums of *cosines* of projections of Gaussian random vectors, and can be thought of as an analogue of the Hanson-Wright inequality (see e.g. [34]) for sums of squares of projections of Gaussian random vectors. It can be viewed as extending the theory of concentration of the original RFF map of Rahimi and Recht [30], as well as the low-dimensional embedding of Chen and Phillips [13] (see Remark 9). Such trigonometric functions of Gaussian random vectors seem to not have been previously investigated in the literature [29, 33], and we call them *Gaussian cosine chaoses* [1]. We believe that the study of cosine chaoses is of independent interest and will find further applications.

Organization of the paper

The rest of this paper is organized as follows. In Section 2 we provide some necessary background and preliminary details. In Section 3 we study the stability of the GKPD weight function over P, under low-distortion dimensionality-reducing maps, and state Lemma 7. Section 4 has the statement of the concentration inequality for Gaussian cosine chaoses, essential for the proof of Lemma 7. In Section 5 we prove some properties of minimum enclosing balls of weighted points in a Hilbert space, and prove the Simplex Distortion Lemma 12. In Section 6 we show how the *Random Fourier Features* map together with Lemma 7 and the Simplex Distortion Lemma give the proof of our main result, Theorem 13. We conclude with a few remarks and open questions in Section 7.

2 Background

We briefly introduce some of the definitions and tools needed for our results and proofs. For a deeper picture, the references [8, 11] would be greatly beneficial to the reader. We also refer the interested reader to [4, 28] for further reading.

2.1 Persistent Homology

Let V be a finite set. An *(abstract) simplicial complex* with vertex set V is a set K of finite subsets of V such that if $A \in K$ and $B \subseteq A$, then $B \in K$. The sets in K are the *simplices* of K. A simplex $F \in K$ that is contained (resp. strictly contained) in a simplex $A \in K$, is said to be a *face* (resp. *proper face*) of A.

A simplicial complex K with a function $f : K \to \mathbb{R}$ such that $f(\sigma) \leq f(\tau)$ whenever σ is a face of τ is a *filtered simplicial complex*. The *sublevel set* of f at $r \in \mathbb{R}$, $f^{-1}(-\infty, r]$, is a subcomplex of K. By considering different values of r, we get a nested sequence of subcomplexes (called a *filtration*) of K, $\emptyset = K^0 \subseteq K^1 \subseteq ... \subseteq K^m = K$, where K^i is the sublevel set at value r_i. The *Čech filtration* associated to a finite set P of points in $\mathbb{R}^D$ plays an important role in Topological Data Analysis. The *Čech complex* $\check{C}_\alpha(P)$ is the set of

[1] In the high-dimensional probability literature [34], a *chaos* is a quadratic function of Gaussian random variables. More generally, polynomials of these variables of degree d are sometimes referred to as chaoses of *order* d [29].

simplices $\sigma \subset P$ such that $\text{rad}(\sigma) \leq \alpha$, where $\text{rad}(\sigma)$ is the radius of the smallest enclosing ball of σ, i.e. $\text{rad}(\sigma) \leq \alpha \Leftrightarrow \exists x \in \mathbb{R}^D,\ \forall p_i \in \sigma,\ \|x - p_i\| \leq \alpha$. When α goes from 0 to $+\infty$, we obtain the Čech filtration of P. $\check{C}_\alpha(P)$ can be equivalently defined as the *nerve* of the closed balls $\overline{B}(p, \alpha)$, centered at the points in P and of radius α:

$$\check{C}_\alpha(P) = \{\sigma \subset P | \cap_{p \in \sigma} \overline{B}(p, \alpha) \neq \emptyset\}.$$

By the *Nerve Lemma* (e.g. [16, 7]), we know that the union of balls $U_\alpha = \cup_{p \in P} \overline{B}(p, \alpha)$, $p \in P$, and $\check{C}_\alpha(P)$ have the same homotopy type. Moreover, since the union of balls of a good sample P of a reasonably regular shape X captures the homotopy type of X, computing the Čech complex of P will provide the homotopy type of X. We also recall that a simpler complex called the *α-complex* of P (see e.g. [14]) also captures the homotopy type of $\bigcup_{p \in P} \overline{B}(p, \alpha)$. Our results will apply to both complexes. Finally, given a Čech complex, the *Vietoris-Rips* or *Rips* complex is the maximal simplicial complex that can be built with the same 1-skeleton. The Rips complex and the corresponding *Rips filtration* are often used in computational topology, as they can be constructed using only pairwise distances.

Persistence Diagrams. Persistent homology is a means to compute and record the changes in the topology of the filtered complexes as the parameter α increases from zero to infinity. Edelsbrunner, Letscher and Zomorodian [15] gave an algorithm to compute the PH, which takes a filtered simplicial complex as input, and outputs a sequence of pairs $(\alpha_{birth}, \alpha_{death})$ of real numbers. Each such pair corresponds to a topological feature, and records the values of α at which the feature appears and disappears, respectively, in the filtration. Thus the topological features of the filtration can be represented using this sequence of pairs, which can be represented either as points in the extended plane $\bar{\mathbb{R}}^2 = (\mathbb{R} \cup \{-\infty, \infty\})^2$, called the *persistence diagram* or as a sequence of barcodes (the *persistence barcode*) (see, e.g., [14]). A pair of persistence diagrams $\mathbb{G}$ and $\mathbb{H}$ corresponding to the filtrations (G_α) and (H_α) respectively, are (β, ζ)-*interleaved*, $(\beta \geq 1)$, if for all α, we have that $G_{\alpha/\beta - \zeta} \subseteq H_\alpha \subseteq G_{\alpha\beta + \zeta}$.

The Persistent Nerve Lemma [12] shows that the PH of the Čech filtration is the same as the homology of the sublevel filtrations of the distance function. The same result also holds for the Delaunay filtration [12].

2.2 Distance to Measure and PH with Power Distances

The most common approach in Topological Data Analysis is to consider the distance function given by the shortest distance to a point in V, i.e. $d_V : \mathbb{R}^D \to \mathbb{R}_+$ is $d_V(x) = \inf_{y \in V} d(x, y)$. (Here V is the finite set from Section 2.1). Given this distance function one can construct the Čech filtration by considering the α-offsets of $d_V(.)$ (i.e. the sublevel sets $\{x \in \mathbb{R}^D \mid d_V(x) \leq \alpha\}$) as unions of balls, and computing the nerve of these unions. To address the significant problem of outliers mentioned earlier in the Introduction, Chazal et al. [8] introduced the notion of *distance to measure* (DTM). In practice, computing the nerve of the α-offsets requires measuring the distance at every point in the space, and so, an approximation to the DTM function is required, which is achieved by considering a finitary version of this distance, called the *k-distance*, which translates to a power distance on the set of k-barycenters of the original point cloud [8, 4]. In general, power distances are often used to approximate unwieldy distance functions for computing the PH. The idea is to approximate the square of the distance to P at a point $x \in \mathbb{R}^D$ by the sum of an easily computable squared distance to a point $p \in P$, together with the square of the *weight* of p: $d_P(x)^2 := d'(x, p)^2 - w(p)$, where $d'(x, p)$ is chosen to be a simpler distance function, easier to compute than $d_P(x)$, and

$w(p) \in \mathbb{R}$ is the *weight* of p, which is set to be the negative of a local approximation of the distance function $d_P(.)$ for points in the neighbourhood of p. In the following paragraphs, we discuss the computation of PH with power distances.

Given a set X and a distance function $d : X \times X \to \mathbb{R}$, the pair (X, d) is a *metric space* if the distance function $d(.,.)$ is reflexive, symmetric and obeys the triangle inequality. Let $\widehat{P}$ be a set of weighted points $\widehat{p} = (p, w(p))$ in a metric space $(\mathcal{M}, d)$. The *power distance* between two weighted points $\widehat{p}, \widehat{q} \in (\mathcal{M}, d)$ is defined as $D(\widehat{p}, \widehat{q}) := d(p,q)^2 - w(p) - w(q)$. Accordingly, we need to extend the definition of the Čech complex to sets of weighted points.

▶ **Definition 1** (Weighted Čech Complex). *Let $\widehat{P} = \{\widehat{p}_1, ..., \widehat{p}_n\}$ be a set of weighted points, where $\widehat{p}_i = (p_i, w_i) \in \mathbb{R}^D \times \mathbb{R}$. The α-Čech complex of $\widehat{P}$, $\check{C}_\alpha(\widehat{P})$, is the set of all simplices σ satisfying $\exists x,\ \forall p_i \in \sigma,\ d(x, p_i)^2 \le w_i + \alpha^2$ which means $\exists x,\ \forall p_i \in \sigma,\ D(x, \widehat{p}_i) \le \alpha^2$. Here $D(x, \widehat{p}_i)$ indicates the power distance between the unweighted point x (i.e. $w(x) = 0$) and the weighted point p.*

In other words, the α-Čech complex of $\widehat{P}$ is the nerve of the closed balls $\overline{B}(p_i, r_i^2 = w_i + \alpha^2)$, centered at the p_i and of squared radius $w_i + \alpha^2$ (if negative, $\overline{B}(p_i, r_i^2)$ is imaginary).

The notions of weighted Čech filtrations and their PH now follow naturally.

In the Euclidean case, we equivalently defined the α-Čech complex as the collection of simplices whose *smallest enclosing balls* have *radius* at most α. We can proceed similarly in the weighted case. Let $\widehat{X} \subseteq \widehat{P}$. We define the *radius of* $\widehat{X}$ as

$$\mathrm{rad}^2(\widehat{X}) \quad = \quad \inf_{x \in \mathbb{R}^D} \max_{\widehat{p}_i \in \widehat{X}} D(x, \widehat{p}_i), \tag{1}$$

and the weighted center or simply the *center* of $\widehat{X}$ as a point, denoted by $c(\widehat{X})$, where this minimum is reached, i.e.

$$c \quad = \quad c(\widehat{X}) = \arg \inf_{x \in \mathbb{R}^D} \max_{\widehat{p}_i \in \widehat{X}} D(x, \widehat{p}_i). \tag{2}$$

Later we shall see the uniqueness of the center and the radius under the above definitions 10.

Analogous to the definition of the Rips complex from the Čech complex in the Euclidean case, we define $\mathrm{VR}_\alpha(\widehat{P})$, the *weighted Rips complex with parameter* α, as the maximal simplicial complex that has the same 1-skeleton as the weighted Čech complex $\check{C}_\alpha(\widehat{P})$, computed using the weighted set of points $\widehat{P}$.

2.3 Kernels and Gaussian Kernel Power Distance

A *kernel* $K : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is a similarity function on points in $\mathbb{R}^D$, such that $K(x, x) = 1$ for all $x \in \mathbb{R}^D$. *Reproducing kernels* are a large class of kernels, having the property that given a reproducing kernel K, there exists a lifting map ϕ to a Hilbert space $\mathcal{H}_K$ such that the kernel function lifts to the inner product on $\mathcal{H}_K$, i.e. for all $x, y \in \mathbb{R}^D$, $K(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}_K}$ (see e.g. Aronszajn [3]). The natural distance function induced by the norm on the Hilbert space $\mathcal{H}_K$ gives a distance using the kernel on $\mathbb{R}^D$, as follows.

$$\begin{aligned} \|\phi(x) - \phi(y)\| &= \sqrt{\langle \phi(x) - \phi(y), \phi(x) - \phi(y) \rangle_{\mathcal{H}_K}} \\ &= \sqrt{K(x,x) + K(y,y) - 2K(x,y)} \quad = \quad \sqrt{2(1 - K(x,y))}, \end{aligned}$$

where the last step follows since $K(x, x) = 1$ for all $x \in \mathbb{R}^D$. For *characteristic kernels*, a slightly smaller subset of reproducing kernels, this distance function is a metric [32]. In this paper, we use the Gaussian kernel, which is a characteristic kernel defined as $K(x, y) = \exp\left(-\|x - y\|^2 / 2\sigma^2\right)$. For $x, y \in \mathbb{R}^D$, the kernel distance $D_K(.,.)$ for the Gaussian kernel is thus, $D_K^2(x, y) := 2(1 - e^{-\|x-y\|^2/2\sigma^2})$.

Kernel Distance to Measure

Let μ be the empirical measure on P defined as $\mu = \dfrac{1}{|P|}\sum_{p\in P}\delta_p$, where δ_p is the Dirac delta measure on P. Using this, given $x \in \mathbb{R}^D$, one can conceive of the kernel distance of the *point mass* δ_x to the *measure* μ, as a function of x, which we denote by $d^K_\mu(x) := \left(\dfrac{1}{|P|}\sum_{p\in P} D^2_K(p,x)\right)^{1/2}$. In [28], Phillips, Wang and Zheng investigated the persistent homology of point sets using $d^K_\mu(.)$, when K is a Gaussian kernel. They showed ([28], Theorems 4.1 and 4.2) that the offsets of P obtained using sublevel sets of $d^K_\mu(x)$ in a given range of thresholds, are homotopically equivalent as long as there is no critical point of $d^K_\mu(x)$ in this range, and are stable under perturbations of the input with bounded Hausdorff distance. Thus the offsets of d^K_μ can be used to estimate the topological properties of the point cloud P.

Gaussian Kernel Power Distance

As mentioned in the Introduction and in Section 2.2, computing the persistent homology using d^K_μ precisely would require computing d^K_μ everywhere in $\mathbb{R}^D$. So in order to avoid this computational expense, Phillips, Wang and Zheng [28] approximated d^K_μ by a power distance using weights on the points in P, as $\min_{p\in P}\left(D^2_K(x,p) - w(p)\right)$, where $w : P \to \mathbb{R}$ is the *Gaussian kernel weight function* at p:

$$w(p) \quad := \quad -D^2_K(\mu,p) = -\left(\frac{1}{|P|}\sum_{y\in P} D^2_K(p,y) - \frac{1}{2|P|^2}\sum_{x,y\in P} D^2_K(x,y)\right). \tag{3}$$

Now we can define weighted points $\hat{p} = (p, w(p))$, $p \in P$, with the weight $w(p)$ being defined as in (3). Let $\widehat{P}$ denote the set of weighted points $\hat{p}$ with $p \in P$. The *Gaussian Kernel Power Distance* (GKPD) between a point $x \in \mathbb{R}^D$ and the pointset P, can be expressed as

$$\begin{aligned} &\min_{p\in P}\left(D^2_K(x,p) - w(p)\right) \\ =\ &\min_{p\in P}\left(D^2_K(x,p) + \frac{1}{|P|}\sum_{y\in P} D^2_K(y,p) - \frac{1}{2|P|^2}\sum_{y,z\in P} D^2_K(y,z)\right). \end{aligned} \tag{4}$$

The GKPD between a pair of weighted points $\hat{p}, \hat{q} \in \widehat{P}$ is defined as

$$D^2_K(\hat{p},\hat{q}) \quad = \quad D^2_K(p,q) - w(p) - w(q). \tag{5}$$

This can be extended to a power distance on pairs of points in $\mathbb{R}^D$ by taking $w(x) = 0$ when $x \in \mathbb{R}^D \setminus \{P\}$. Thus the GKPD between a point x and the pointset P (4) can be concisely expressed as $\min_{p\in P}\left(D^2_K(\hat{x},\hat{p})\right)$.

Since the Gaussian kernel distance is a radial function of the Euclidean distance, the level sets of the kernel power distance are also unions of balls. Moreover Phillips, Wang and Zheng [28][Theorem 3.1, Lemma 3.1] showed that up to constant factors, the GKPD approximates the Gaussian power distance to the uniform measure d^K_μ, i.e. for any $x \in \mathbb{R}^D$, the GKPD between x and P, is at least $d^K_\mu(x)^2/2$ and at most $2d^K_\mu(x)^2 + 3D^2_K(p,x)$, where $p \in P$ is the point that achieves the minimum in (4).

▶ Remark 2. Note that the weight $w(p)$ of a point $p \in P$ depends not only on p, but on the pairwise squared Gaussian kernel distances between points in P. This fact introduces a crucial requirement for any embedding or dimensionality reduction procedure: the kernel weight function needs to be recomputed in the image space – since otherwise for each point $p \in P$, we could set $w(f(p)) = w(p)$.

2.4 Random Fourier Features

For points in $\mathbb{R}^D$, there exists a mapping to $\mathbb{R}^{2t}$, with $t = O\left(\varepsilon^{-2}\log n\right)$ that gives a relative approximation of the kernel distance on $\mathbb{R}^D$, as the natural inner product on $\mathbb{R}^{2t}$. This is the well-known Random Fourier Features (RFF) map of Rahimi and Recht [30], which was shown by Chen and Phillips in [13] to give a relative approximation of Gaussian kernel distances in $\mathbb{R}^D$ by Euclidean distances in $\mathbb{R}^{2t}$. Their mapping is given as follows: For $i = 1, \ldots, t$, given $\sigma \geq 0$, let $\omega_i \sim \mathcal{N}_D(0, \sigma^{-2})$ be independent D-dimensional Gaussian vectors, and define the random map $f_i : \mathbb{R}^D \to \mathbb{R}^2$, as

$$f_i(x) \quad = \quad (\cos(\langle \omega_i, x\rangle), \sin(\langle \omega_i, x\rangle)). \tag{6}$$

Finally, define the mapping $f : \mathbb{R}^D \to \mathbb{R}^{2t}$ as

$$f(x) \quad = \quad \frac{1}{\sqrt{t}}\left(\bigoplus_{i=1}^{t} f_i(x)\right). \tag{7}$$

▶ **Theorem 3** (Chen, Phillips [13]). *Given any $\varepsilon, \delta \in (0,1)$, for any set $P \subset \mathbb{R}^D$ of n points, the RFF map $f : \mathbb{R}^D \to \mathbb{R}^{2t}$ as defined above with $t := \Omega\left(\varepsilon^{-2}\log(n/\delta)\right)$ dimensions is an ε-distortion map for the Gaussian kernel distance, i.e. f satisfies $\frac{\|f(x)-f(y)\|^2}{D_K^2(x,y)} \in (1-\varepsilon, 1+\varepsilon)$ for all pairs of points $x, y \in P$, with probability at least $1-\delta$.*

Moreover, the RFF map $f : \mathbb{R}^D \to \mathbb{R}^{2t}$ can be computed in time $O(nt)$.

▶ Remark 4. Recall from Section 2.3 that for the Gaussian kernel, there exists a lifting map to a Hilbert space $\mathcal{H}_K$ such that the RFF lifts to the inner product on $\mathcal{H}_K$. Theorem 3 shows that for any given finite set of points, the inner product on $\mathcal{H}_K$ (infinite-dimensional) can be approximated by the Euclidean inner product on a finite-dimensional space.

3 Low-Distortion Maps for Power Distances

In this section, we shall look at low-distortion mappings of power distances. First we need the notion of an *ε-distortion* map for power distances between metric spaces.

▶ **Definition 5.** *Given metric spaces (X, d_X) and (Y, d_Y), a point set $P \subset X$, $\eta \geq 0$, and $\varepsilon \in (0,1)$, a mapping $f : X \to Y$ is an* ε-distortion map with additive factor η, *or an* (ε, η)-distortion map, *between pairwise d_X-distances in P and d_Y- distances in $f(P)$, if*

$$\forall x, y \in P: \;\; (1-\varepsilon)d_X(x,y)^2 - \eta \leq d_Y(f(x), f(y))^2 \leq (1+\varepsilon)d_X(x,y)^2 + \eta.$$

Further, given a pair of weight functions $w_X : P \to \mathbb{R}$ and $w_Y : P \to \mathbb{R}$, f is an (ε, η)-distortion map *between w_X and w_Y, if, $\forall x \in P: \;\; |w_Y(f(x)) - w_X(x)| \leq \varepsilon|w_X(x)| + \eta$.*

$$\forall x \in P: \;\; |w_Y(f(x)) - w_X(x)| \leq \varepsilon|w_X(x)|.$$

The definition of an (ε, η) distortion map now extends naturally to the case of the GKPD.

▶ **Definition 6.** *Given $D, t > 0$ and the power distance D_K^2 defined as in (5), the mapping $f : \widehat{\mathbb{R}^D} \to \widehat{\mathbb{R}^{2t}}$ is an* (ε, η)-distortion mapping *between D_K^2 in $\widehat{\mathbb{R}^D}$ and Euclidean distances in $\widehat{\mathbb{R}^{2t}}$, if f is an (ε, η)-distortion mapping between the unweighted Gaussian kernel distance D_K^2 in $\mathbb{R}^D$ and Euclidean distances in $\mathbb{R}^{2t}$, and is also an (ε, η)-distortion mapping for the weight function $w(.)$ defined in (3).*

In the context of Gaussian kernels, perhaps the most well-known example of an $(\varepsilon, 0)$-distortion map is the RFF map, as applied by Chen and Phillips in [13].

Let f be an (ε, η)-distortion mapping for some $\varepsilon, \eta \in \mathbb{R}_+$. From Remark 2, we know that the Gaussian weight function needs to be recomputed in the image space. For each $p \in P$, let us define

$$w(f(p)) := -\left(\frac{1}{|P|}\sum_{y\in P}\|f(p)-f(y)\|^2 - \frac{1}{2|P|^2}\sum_{x,y\in P}\|f(x)-f(y)\|^2\right). \tag{8}$$

Now we have a natural extension of f which acts on the weighted points $\hat{p}$ – we define $f(\widehat{p}) := (f(p), w(f(p)))$. This notion also extends to sets of weighted points. Thus for a set Σ of points, we use $f(\widehat{\Sigma})$ to mean the set $\{f(\hat{s}) | s \in \Sigma\}$.

We now turn to the distortion bound for the weights $w(p)$ under the map f. Let us define $T_1(p) := \frac{1}{|P|}\sum_{y\in P\setminus\{p\}} D_K^2(y,p)$, and $T_2 := \frac{1}{2|P|^2}\sum_{x\neq y;\, x,y\in P} D_K^2(x,y)$. Then $-w(p) = T_1(p) - T_2$. Observe that both $T_1(p)$ and T_2 are averages of GKPDs, and thus, $-w(p)$ is a *difference* of mean GKPDs. Therefore, in order to bound the distortion of $-w(p)$, it is not enough to have an ε-distortion map since, if both $T_1(p)$ and T_2 are ε-distorted, then $w(f(p)) - w(p) \le ((1+\varepsilon)T_1(p) - T_1(p)) - ((1-\varepsilon)T_2 - T_2)$, which can be as large as $\varepsilon(T_1(p) + T_2)$. Note that the ratio $\frac{T_1(p)+T_2}{T_1(p)-T_2}$ could be arbitrarily large, as $T_1(p)$ can be arbitrarily close to T_2. Therefore, we need a stronger condition on the distortion of $T_1(p)$ and T_2 to ensure that the distortion of $-w(p)$, is at most an additive factor of $\varepsilon(T_1(p) - T_2)$. This is where the following lemma comes to our aid. Before stating the lemma, we recall the definition of the stable rank of a matrix.

The *stable rank* of a matrix M is the ratio i.e. $r_{st}(M) := \|M\|_F^2/\|M\|^2$, where $\|M\|_F$ is the Frobenius norm of M, (given by $(\sum_{i,j} M_{ij}^2)^{1/2}$) and $\|M\|$ is the operator norm of M ($\max_{\|x\|=1}\|Mx\|$), see e.g. [34] [Chapter 6]. Given a set S of vectors, let r_{st} denote the stable rank of the $D \times s$ matrix $\tilde{S}$, whose columns are the vectors $v \in S$. For each $p \in P$, define $S(p) := \{p - y \mid y \in P \setminus \{p\}\}$. Further, define $S_2 := \{x - y \mid x \neq y \in P\}$, and let $\tilde{S}(p)$, $\tilde{S_2}$ denote the corresponding matrices.

▶ **Lemma 7** (Distortion of Weights). *Given $\delta \in (0,1]$, and a function $\gamma = \gamma(n)$ such that for each $p \in P$, $r_{st}(\tilde{S}(p)) \ge \gamma(n)$ and $r_{st}(\tilde{S_2}) \ge \gamma(n)$, then with probability at least $1-\delta$, for each $p \in P$, the RFF map (7) with $t \ge C\frac{\log n/\delta}{\varepsilon^2}$ satisfies the following*

$$(1-\varepsilon)w(p) - (\gamma(n))^{-1/2} \le w(f(p)) \le (1+\varepsilon)w(p) + (\gamma(n))^{-1/2}.$$

Utilizing the fact that both $T_1(p)$ and T_2 are *averages* of squared distances, Lemma 7 shows that under certain conditions, we can get a stronger concentration of $T_1(p)$, for each $p \in S$, as well as for T_2. This will allow us to bound the distortion for the weights $w(p)$, $p \in P$, by a $o(1)$ additive error.

4 Concentration Inequality for Gaussian Cosine Chaos

In this section, we shall state a concentration inequality for certain trigonometric functions of projections of Gaussian random vectors, which will be crucially used in our bound on the distortion of the GKPD weight function under the RFF map. The inequality is described in the following general framework. Let $S = \{v_1, \ldots, v_s\}$ be a set of s vectors in $\mathbb{R}^D$. For $k = 1, \ldots, t$, let g_k be independent and identically distributed standard normal vectors in $\mathbb{R}^D$, and define

$$L_t = L_t(S) \quad := \quad \frac{1}{4s}\sum_{v\in S}\frac{1}{t}\sum_{k=1}^{t}(1-\cos(\langle v, g_k\rangle)).$$

As we shall see in the proof of Lemma 7, under the RFF map, the sum of squares of Gaussian kernel distances of a given set of pairs of points has the same distribution as a random variable of the form L_t. Thus, a concentration bound for L_t shall allow us to bound the distortions of $T_1(p)$, for every $p \in P$ as well as T_2, up to a $(1 \pm o(1))$-factor (under the stable rank assumption). Here $T_1(p)$ and T_2 are as defined in Section 2.3.

To understand the need for a new inequality, let us consider the analogous situation when we have a set of vectors S and t standard Gaussian vectors $g_i \in \mathcal{N}(0, I_D)$, $i \in [t]$, and we are interested in concentration bounds for the function $Q(S) := \sum_{v \in S} \sum_{i=1}^{t} \langle v, g_i \rangle^2 = \|G^\top V\|_F^2$, where V is the matrix whose columns are $v \in S$, and G has column vectors g_i, $i \in [t]$. $Q(S)$ is a quadratic function of Gaussian random vectors, often referred to as a *Gaussian chaos* [34]. In this case, the Hanson-Wright inequality [34][Chapter 6] could be used to obtain a strong concentration bound in terms of the stable rank of V, allowing us to exploit any mutual orthogonality present between the vectors $v \in S$. Using the Taylor series expansion for $\cos(x)$, L_t can be rewritten as $L_t = \sum_{i=1}^{t} \sum_{j \geq 1} \frac{(-1)^{j-1}}{(2j)!} \|V^\top g_i\|_{2j}^{2j}$. Thus $L_t(S)$ – which we call a *Gaussian cosine chaos* – is a linear combination of infinitely many Gaussian chaoses of increasing order, and therefore cannot be addressed by existing concentration inequalities for Gaussian chaoses of bounded order (see e.g. Latała [29] and Talagrand [33] for recent results on chaoses.) We are not aware of the existence of any such inequality for chaoses of unbounded order, prior to our result.

Let $u_j \in \mathbb{R}^D$, $j = 1, \ldots, r$ be an orthonormal basis for the span of $v_1, \ldots, v_s$. For each $i, i' = 1, \ldots, s$, define $w_{ii'} \in \mathbb{R}^r$ as $(w_{ii'})_j = |\langle v_i, u_j \rangle| - |\langle v_{i'}, u_j \rangle|$, $j \in [r]$. Finally, recall that $r_{st} := r_{st}(\tilde{S})$ is the stable rank of the matrix $\tilde{S}$ having column vectors $v_1, \ldots, v_s$. We obtain the following concentration inequality for L_t. (For the proof, we refer the reader to the complete version).

▶ **Theorem 8** (Concentration inequality for Gaussian cosine chaos). *For any $\varepsilon \in [0, 1]$, the following holds.*

1. *For any set S of vectors in $\mathbb{R}^D$,*
$$\mathbb{P}\left[|L_t - \mathbb{E}[L_t]| \geq \varepsilon\,\mathbb{E}[L_t]\right] \leq 2 \cdot \exp\left(-\frac{\varepsilon^2 t\,\mathbb{E}[L_t]^2}{3r \sum_{1 \leq i < i' \leq s} e^{-\|w_{ii'}\|^2/(2r)}}\right). \quad (9)$$
2. *If every vector $v \in S$ has Euclidean norm at most 1, then there exists $C > 0$ such that*
$$\mathbb{P}\left[|L_t - \mathbb{E}[L_t]| \geq \varepsilon\,\mathbb{E}[L_t]\right] \leq 2 \cdot \exp\left(-\left(C\varepsilon^2 t r_{st}\right)\right). \quad (10)$$

▶ Remark 9. The above inequality generalizes and extends the concentration inequality of Chen and Phillips [13] (Appendix) for RFF maps. Further, since the square of the Gaussian kernel distance is a norm obtained using the kernel density estimate as an inner product, Theorem 8 may also be viewed as a generalization of the concentration inequality of Rahimi and Recht [30] for the RFF map. For the case when the vectors in S have bounded norm, Theorem 8 improves with the stable rank r_{st}. In the worst case, when $r_{st} = 1$, we get back the Chen-Phillips inequality.

To prove Theorem 8, we implicitly build a Doob martingale on the random variable, by sequentially exposing the coordinates of the Gaussian random vector with respect to an orthonormal basis for the subspace spanned by them. (Typically we shall choose the basis given by the left singular vectors of $\tilde{S}$, where $\tilde{S}$ is as defined before Lemma 7.) Theorem 8 is then proved by computing explicit bounds for the path variance of this martingale. In the case when the vectors have bounded Euclidean norm, we use the Singular Value Decomposition (SVD) of the matrix $\tilde{S}$, to bound the path variance in terms of the stable rank of $\tilde{S}$, getting a condition similar to the Hanson-Wright inequality.

5 Minimum Enclosing Power Balls

As mentioned earlier, the weighted Čech complex can be easily constructed once we know the minimum enclosing ball of subsets of weighted points. We thus need to show that such balls are almost preserved under our dimensionality reduction procedure. This will be obtained via the main result of this section – a decomposition theorem for the squared radius of the minimum enclosing ball of a set of weighted points under the GKPD, in terms of a linear combination of pairwise GKPDs of the weighted points. To prove such a decomposition, we need to understand some properties of minimum enclosing balls of collections of weighted points under power distances, in particular the GKPD. These properties will be central in the proof of our main result, proved in the next section.

The primary challenge in proving such a result comes from the fact that squared Gaussian kernel distances are non-linear, so that techniques used for squared Euclidean distances do not apply. To address, we need the crucial fact that there exists a lifting map from $\mathbb{R}^D$ to an (infinite dimensional) Hilbert space $\mathcal{H}_K$, which maps RFFs to the inner product on $\mathcal{H}_K$.

Consider a set of points $p_1, \ldots, p_k \in \mathbb{R}^D$ weighted using the GKPD as defined in (3), and let $\widehat{\sigma}$ denote the associated abstract simplex formed by $\{\widehat{p}_1, \ldots, \widehat{p}_k\}$. Using the map to the Hilbert space $\mathcal{H}_K$ discussed in Section 2.3, which lifts the Gaussian kernel function to the inner product on $\mathcal{H}_K$, together with convexity and perturbation arguments, we prove properties of the minimum enclosing balls of subsets of weighted points, such as the uniqueness of their centre and the radius. These properties – summarised in Proposition 10 – will be useful in obtaining our distortion bound on the radius of the minimum enclosing ball. (For their proofs, we refer the reader to the complete version.)

▶ **Proposition 10.**

1. *For the simplex $\widehat{\sigma}$, its center $c(\widehat{\sigma})$ and radius $\mathrm{rad}(\widehat{\sigma})$ are unique.*
2. *There exists a set of non-negative reals $(\lambda_i)_{i\in[k]}$, such that $\sum_{i\in[k]} \lambda_i = 1$, $\sum_{i\in[k]} \lambda_i p_i = c$, and*

$$\mathrm{rad}^2(\widehat{\sigma}) = \frac{1}{2} \sum_{i\in[k]} \sum_{j\in[k]} \lambda_i \lambda_j D_K^2(\widehat{p}_i, \widehat{p}_j).$$

Further, $\lambda_i = 0$ for all $i \in [k]$ such that $D_K(c, \widehat{p}_i) < \mathrm{rad}^2(\widehat{\sigma})$.

From the above proposition the following Decomposition Lemma can be proved, which shows that the squared radius of the minimum enclosing ball of the weighted point set $\widehat{X}$ can be expressed as a combination of pairwise power distances of the points in $\widehat{X}$.

Let I be the set of indices of $\widehat{p_j} \in \hat{X}$, such that $\mathrm{rad}^2(\widehat{X}) = D_K(c, \hat{p_j})$. Let λ_i, $i \in I$ be such that $\sum_{i\in I} \lambda_i p_i = c$, where c is the center of $\widehat{\sigma}$.

▶ **Lemma 11** (Decomposition Lemma). *Let I be the set of indices as defined above, and let $(\lambda_i)_{i\in I}$ be the corresponding set of non-negative reals, as defined above. Then*

$$\mathrm{rad}^2(\widehat{X}) = \frac{1}{2} \sum_{i\in I} \sum_{j\in I} \lambda_i \lambda_j D_K(\widehat{p}_i, \widehat{p}_j).$$

As a direct consequence of Proposition 10 together with the Decomposition Lemma 11, we obtain the Simplex Distortion Lemma, which bounds the distortion of the radius of the minimum enclosing ball of each simplex under the Čech filtration.

▶ **Lemma 12** (Simplex Distortion Lemma). *Let $\widehat{\sigma} \subset \widehat{P}$ be a simplex in the weighted Čech complex $\check{C}_\alpha(\widehat{P})$ using the GKPD $D_K^2(\hat{p}, \hat{q})$ defined in (5), where $\hat{p}, \hat{q} \in \widehat{P}$ and the weights are defined in (3) and let $f : (\mathbb{R}^D, D_K) \to (\mathbb{R}^{2t}, \|.\|)$ be an (ε, η)-distortion map for the* *pairwise GKPD Then with $f(\widehat{\sigma})$ denoting the image of the simplex σ in $\mathbb{R}^{2t}$, with the weights recomputed using* (8),

$$(1-\varepsilon)(\mathrm{rad}^2(\widehat{\sigma})-\eta) \leq \mathrm{rad}^2(f(\widehat{\sigma})) \leq (1+\varepsilon)(\mathrm{rad}^2(\widehat{\sigma})+\eta),$$

Proof of Simplex Distortion Lemma 12. Let the simplex $\widehat{\sigma} = \{\widehat{p}_1, \ldots, \widehat{p}_k\}$, where for all $i \in [k]$, $\widehat{p}_i := (p_i, w(p_i))$ is a weighted point, and let $c(\widehat{\sigma})$ and $\mathrm{rad}(\widehat{\sigma})$ denote its center and radius respectively.

Since the Gaussian kernel is a characteristic kernel, there exists a Hilbert space $\mathcal{H}_K$, and a lifting map $\phi : \mathbb{R}^D \to \mathcal{H}_K$, such that for all $x, y \in \mathbb{R}^D$, $D_K(x,y) = \|\phi(x) - \phi(y)\|_{\mathcal{H}_K}$ (see e.g. [13, 27]). By eqn. (3) the weight of a point $p \in P$ is a weighted sum of squared kernel distances:

$$w(p) := -D_K^2(\mu, p) = -\left(\frac{1}{|P|}\sum_{y\in P} D_K^2(p,y) - \frac{1}{2|P|^2}\sum_{x,y\in P} D_K^2(x,y)\right).$$

Thus the lifting map ϕ extends naturally to the weights $w(p_i)$, $i \in [k]$, as

$$\phi(w(p_i)) = -\left(\frac{1}{|P|}\sum_{y\in P}\|\phi(p)-\phi(y)\|^2_{\mathcal{H}_K} - \frac{1}{2|P|^2}\sum_{x,y\in P}\|\phi(x)-\phi(y)\|^2_{\mathcal{H}_K}\right),$$

which allows us to define the weights in the lifted space as $w(\phi(p)) := \phi(w(p))$.

Applying the Decomposition Lemma 11 with $\mathcal{H} = \mathcal{H}_K$, and the lifted weighted points given by $\phi(\hat{p}_i) := (\phi(p_i), \phi(w(p_i)))$, we have

$$\mathrm{rad}^2(\widehat{\sigma}) \quad = \quad \frac{1}{2}\sum_{i,j\in[k]} \lambda_i\lambda_j D_K^2(\widehat{p}_i, \widehat{p}_j). \tag{11}$$

Since f is an (ε,η)-distortion map, for each pair $\widehat{p}_i, \widehat{p}_j \in \widehat{\sigma}$, we have

$$(1-\varepsilon)\|\phi(p_i)-\phi(p_j)\|^2_{\mathcal{H}_K} \quad\leq\quad \|f(p_i)-f(p_j)\|^2 \quad\leq\quad (1+\varepsilon)\|\phi(p_i)-\phi(p_j)\|^2_{\mathcal{H}_K} \tag{12}$$

Since f is an (ε,η)-distortion map for the weight function, we get for each $\widehat{p}_i \in \widehat{\sigma}$,

$$(1-\varepsilon)w(\widehat{p}_i) - \eta \quad\leq\quad w(f(\widehat{p}_i)) \quad\leq\quad (1+\varepsilon)w(\widehat{p}_i) + \eta. \tag{13}$$

Subtracting the weights $w(f(\widehat{p}_i)), w(f(\widehat{p}_j))$ from the squared distance $\|f(p_i)-f(p_j)\|^2$, and using that $D_K^2(\hat{p}_i, \hat{p}_j) = D_K^2(p_i,p_j) - w(p_i) - w(p_j) = \|\phi(p_i)-\phi(p_j)\|^2_{\mathcal{H}_K} - w(p_i) - w(p_j)$, we get

$$(1-\varepsilon)D_K^2(\widehat{p}_i,\widehat{p}_j) - 2\eta \quad\leq\quad D_K^2(f(\widehat{p}_i), f(\widehat{p}_j)) \quad\leq\quad (1+\varepsilon)D_K^2(\widehat{p}_i,\widehat{p}_j) + 2\eta. \tag{14}$$

Let $f(\widehat{\sigma})$ denote the image of the simplex $\widehat{\sigma}$ under the map f, and $c(f(\widehat{\sigma}))$ be its center. Applying Proposition 10(i) on the space $(\mathbb{R}^{2t}, \|.\|)$, we get that $c(f(\widehat{\sigma}))$ is a convex combination of the vertices of $f(\widehat{\sigma})$, say

$$c(f(\widehat{\sigma})) = \sum_{i\in[k]} \mu_i f(\widehat{p}_i),$$

where $\forall i \in [k]$; $\mu_i \geq 0$ and $\sum_{i\in[k]} \mu_i = 1$[2]. Since f is an (ε, η)-distortion map, using the Decomposition Lemma 11 we get

$$\begin{aligned}\mathrm{rad}^2(f(\widehat{\sigma})) &= \sum_{i,j\in[k]} \mu_i\mu_j \left(\frac{1}{2}D_K^2(f(\widehat{p_i}), f(\widehat{p_j}))\right) \\ &\geq \sum_{i,j\in[k]} \frac{\mu_i\mu_j}{2}\left((1-\varepsilon)D_K^2(\widehat{p_i},\widehat{p_j}) - 2\eta\right) \qquad (15)\end{aligned}$$

$$\text{i.e.} \sum_{i,j\in[k]} \frac{\mu_i\mu_j}{2}\left(D_K^2(\widehat{p_i},\widehat{p_j}) - \eta\right) \leq \frac{\mathrm{rad}^2(f(\widehat{\sigma}))}{1-\varepsilon}. \qquad (16)$$

Also, by the minimality in the definition of the squared radius of a weighted simplex, we have

$$\mathrm{rad}^2(\widehat{\sigma}) = \frac{1}{2}\sum_{i,j\in[k]} \lambda_i\lambda_j D_K^2(\widehat{p_i},\widehat{p_j}) \leq \frac{1}{2}\sum_{i,j\in[k]} \mu_i\mu_j D_K^2(\widehat{p_i},\widehat{p_j}), \quad \text{and} \qquad (17)$$

$$\begin{aligned}\mathrm{rad}^2(f(\widehat{\sigma})) &= \frac{1}{2}\sum_{i,j\in[k]} \mu_i\mu_j D_K^2(f(\widehat{p_i}), f(\widehat{p_j})) \leq \frac{1}{2}\sum_{i,j\in[k]} \lambda_i\lambda_j D_K^2(f(\widehat{p_i}), f(\widehat{p_j})), \\ &\leq \sum_{i,j\in[k]} \frac{\lambda_i\lambda_j}{2}\left((1+\varepsilon)D_K^2(\widehat{p_i},\widehat{p_j}) + 2\eta\right), \qquad (18) \\ &= (1+\varepsilon)(\mathrm{rad}^2(\widehat{\sigma}) + \eta) \qquad (19)\end{aligned}$$

where in step (18) we again used that G is an ε-distortion map, and the last step followed from the Decomposition Lemma 11. Combining equations (16), (17) and (19) gives

$$(1-\varepsilon)(\mathrm{rad}^2(\widehat{\sigma}) - \eta) \leq \mathrm{rad}^2(f(\widehat{\sigma})) \leq (1+\varepsilon)(\mathrm{rad}^2(\widehat{\sigma}) + \eta), \qquad (20)$$

which completes the proof of the lemma. ◀

6 Main Result

In this section we shall prove the following theorem, which is our main result. Recall that for each $p \in P$, $\tilde{S}(p)$ is the matrix with column vectors $x - p$, $x \in P \setminus \{p\}$, and that $\tilde{S}_2$ is the matrix with column vectors $x - y$, where $x, y \in P$ and $x \neq y$. Further, recall that $\check{C}_\alpha(\widehat{P})$ is the Čech filtration computed on $\widehat{P}$ using the GKPD and $\check{C}_\alpha(f(\widehat{P}))$ is the Čech filtration computed using Euclidean distances between the images of the weighted points $\widehat{P}$ under the RFF map $f : \mathbb{R}^D \to \mathbb{R}^{2t}$, and the Rips filtrations $\mathrm{VR}_\alpha(\widehat{P})$ and $\mathrm{VR}_\alpha(f(\widehat{P}))$ are defined similarly.

▶ **Theorem 13.** *Given $\sigma > 0$, $\varepsilon, \delta_0 \in (0,1)$, $K > 0$, a finite set $P \subset B(0,K) \subset \mathbb{R}^D$ consisting of n points, then with probability at least $1 - \delta_0$, a Random Fourier Features projection map $f : \mathbb{R}^D \to \mathbb{R}^{2t}$ onto $2t := \Omega\left(\varepsilon^{-2}\log(n/\delta)\right)$ dimensions is such that $\check{C}_\alpha(f(\widehat{P}))$ is $((1-\varepsilon)^{-1/2}, \gamma^{-1/2})$-interleaved with $\check{C}_\alpha(\widehat{P})$ provided that the matrices $\tilde{S}(p)$, $p \in S$, and $\tilde{S}_2$ have stable rank at least $\gamma(n)$, where $\lim_{n\to\infty}\gamma(n) = \infty$. Further, the corresponding Rips filtrations $VR_\alpha(\widehat{P})$ and $VR_\alpha(f(\widehat{P}))$ are also $((1-\varepsilon)^{-1/2}, \gamma^{-1/2})$-interleaved.*

[2] Note that the convex combination $(\mu_i)_{i\in[k]}$ need not be the same as the combination $(\lambda_i)_{i\in[k]}$ for $c(\widehat{\sigma})$) in the original space.

Proof of Theorem 13. In order to prove that an RFF mapping onto $2t$ dimensions gives a data set whose weighted Čech filtration (with the weights being recomputed in the image space) is interleaved with the original filtration, it suffices to show that with high probability, for an arbitrary weighted simplex σ the radius of σ under the GKPD is $(1 \pm \varepsilon)$-distorted with an additive factor which is $o(1)$, under the RFF mapping. We set $t \geq C'\varepsilon^{-2}\log n$, choosing C' sufficiently large, and construct the RFF map $f : \mathbb{R}^D \to \mathbb{R}^{2t}$ defined as in (7), By applying Theorem 3 and Lemma 7 setting $\delta = \delta_0/2$ in each case, we get that the statements (i) $\forall x, y \in P: \;\; (1-\varepsilon)D_K^2(x,y) \leq \|f(x) - f(y)\|^2 \leq (1+\varepsilon)D_K^2(x,y)$, and (ii) $\forall x \in P: \;\; (1-\varepsilon)w(f(x)) - 1/\gamma \leq \|f(x) - f(y)\|^2 \leq (1+\varepsilon)w(f(x)) + 1/\gamma$, each hold with probability at least $1 - \delta_0/2$, and therefore, they hold simultaneously with probability at least $1 - \delta_0$. Thus with probability at least $1 - \delta_0$, f is an $(\varepsilon, 1/\gamma)$-distortion map for the kernel distance D_K.

This immediately implies the statement for the Rips filtration, since the weighted Rips filtration on $\widehat{P}$ is defined using only pairwise distances and pointwise weights.

For the Čech filtration, we apply the Simplex Distortion Lemma 12 with the mapping f obtained above and $t \geq C\varepsilon^{-2}\log n$ for a sufficiently large constant C, to get that for *each* simplex in $\check{C}_\alpha(f(\widehat{P}))$, the square of its radius is distorted by at most a multiplicative factor of $(1 \pm \varepsilon)$, together with a $o(1)$ additive factor. Therefore, the weighted Čech filtration $\check{C}_\alpha(\widehat{P})$ built using the kernel distance function D_K interleaves with $\check{C}_\alpha(f(\widehat{P}))$, i.e. the Čech filtration built on the image of the weighted point set $\widehat{P}$ under f using the Euclidean distance, as follows, $\check{C}_{\alpha_-}(\widehat{P}) \subseteq \check{C}_\alpha(f(\widehat{P})) \subseteq \check{C}_{\alpha_+}(\widehat{P})$. where $\alpha_- := \alpha\sqrt{(1-\varepsilon) - 1/\gamma(n)}$ and $\alpha_+ := \alpha\sqrt{(1+\varepsilon) + 1/\gamma(n)}$. Together with the fact that $1 + \varepsilon \leq (1-\varepsilon)^{-1}$ for $\varepsilon \in (0,1)$, this completes the proof for the Čech filtration. ◀

7 Conclusion

We have shown that the Random Fourier Features map can be used to reduce the dimensionality of input data, for building persistence diagrams using the Čech filtration. Our results also apply to the weighted Rips filtration constructed using the GKPD. Further, since the Čech and the Delaunay complexes are nerves of good coverings of the same union of balls, it follows that our results also hold for the Delaunay filtration.

The computational complexity of our embedding may be bounded by observing that it involves a one-time pre-multiplication of the data matrix by a $D \times t$ Gaussian matrix; thus a naive implementation would require $O(nDt)$ operations. Further, computing the weights of the data points involves another $O(n^2)$ operations, so that the total time complexity comes to $O(n(n + Dt))$. While this may seem expensive, note that it is just a preprocessing step, and the computational cost is more than offset by the subsequent ease of working with Euclidean distances instead of Gaussian kernel distances.

Our embedding works well for datasets with high stable rank. This is complementary to the zone of operability of traditional dimensionality reduction techniques such as Principal Component Analysis (PCA), which work well when most of the energy of the data vectors is concentrated in a few principal directions. By combining PCA for the principal components of the data matrix with our embedding for the other components, it may be possible to obtain hybrid low-dimensional embeddings which would work well in practice with both high and low stable rank datasets. On the other hand, PCA can widely distort individual pairwise distances, and therefore often does not yield theoretical guarantees on the distortion of the output.

An interesting application of our technique could therefore be to combine it with other dimensionality reduction techniques such as PCA or gradient descent based techniques, to obtain mixed dimensionality reduction schemes which would work independent of stable rank assumptions.

Finally, it would also be interesting to prove similar dimensionality reduction results for PH computed using other classes of characteristic kernels, or even more general kernels.

References

1 Thomas D. Ahle, Michael Kapralov, Jakob Bæk Tejs Knudsen, Rasmus Pagh, Ameya Velingker, David P. Woodruff, and Amir Zandieh. Oblivious sketching of high-degree polynomial kernels. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 141–160. SIAM, 2020. doi:10.1137/1.9781611975994.9.

2 Hirokazu Anai, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hiroya Inakoshi, Raphaël Tinarrage, and Yuhei Umeda. DTM-Based Filtrations. In Gill Barequet and Yusu Wang, editors, *35th International Symposium on Computational Geometry, SoCG 2019, June 18-21, 2019, Portland, Oregon, USA*, volume 129 of *LIPIcs*, pages 58:1–58:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.SoCG.2019.58.

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

4 Shreya Arya, Jean-Daniel Boissonnat, Kunal Dutta, and Martin Lotz. Dimensionality reduction for k-distance applied to persistent homology. *J. Appl. Comput. Topol.*, 5(4):671–691, 2021. doi:10.1007/s41468-021-00079-x.

5 Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, and Amir Zandieh. Random fourier features for kernel ridge regression: Approximation bounds and statistical guarantees. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pages 253–262. JMLR.org, 2017.

6 Yair Bartal, Ben Recht, and Leonard J. Schulman. Dimensionality reduction: Beyond the Johnson-Lindenstrauss bound. In Dana Randall, editor, *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, pages 868–887. SIAM, 2011. doi:10.1137/1.9781611973082.68.

7 Karol Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fundamenta Mathematicae*, 35(1):217–234, 1948. URL: http://eudml.org/doc/213158.

8 Mickaël Buchet, Frédéric Chazal, Steve Y. Oudot, and Donald R. Sheehy. Efficient and robust persistent homology for measures. *Comput. Geom.*, 58:70–96, 2016. doi:10.1016/j.comgeo.2016.07.001.

9 Mickaël Buchet, Tamal K. Dey, Jiayuan Wang, and Yusu Wang. Declutter and resample: Towards parameter free denoising. *JoCG*, 9(2):21–46, 2018. doi:10.20382/jocg.v9i2a3.

10 Moses Charikar and Paris Siminelakis. Hashing-based-estimators for kernel density in high dimensions. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1032–1043, 2017. doi:10.1109/FOCS.2017.99.

11 Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Found. Comput. Math.*, 11(6):733–751, 2011. doi:10.1007/s10208-011-9098-0.

12 Frédéric Chazal and Steve Oudot. Towards persistence-based reconstruction in Euclidean spaces. In Monique Teillaud, editor, *Proceedings of the 24th ACM Symposium on Computational Geometry, College Park, MD, USA, June 9-11, 2008*, pages 232–241. ACM, 2008. doi:10.1145/1377676.1377719.

13 Di Chen and Jeff M. Phillips. Relative Error Embeddings of the Gaussian Kernel Distance. In *International Conference on Algorithmic Learning Theory, ALT 2017, 15-17 October 2017, Kyoto University, Kyoto, Japan*, pages 560–576, 2017. URL: http://proceedings.mlr.press/v76/chen17a.html.

14 Herbert Edelsbrunner and John Harer. *Computational Topology - an Introduction*. American Mathematical Society, 2010. URL: `http://www.ams.org/bookstore-getitem/item=MBK-69`.

15 Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discret. Comput. Geom.*, 28(4):511–533, 2002. `doi:10.1007/s00454-002-2885-2`.

16 Robert Ghrist. *Elementary Applied Topology*. Createspace, September 2014. URL: `https://www2.math.upenn.edu/~ghrist/notes.html`.

17 C. Giraud. *Introduction to High-Dimensional Statistics*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2014.

18 Leonidas J. Guibas, Dmitriy Morozov, and Quentin Mérigot. Witnessed k-distance. *Discret. Comput. Geom.*, 49(1):22–45, 2013. `doi:10.1007/s00454-012-9465-x`.

19 Thomas Hofmann, Bernhard Schölkopf, and Alexander J. Smola. Kernel methods in machine learning. *The Annals of Statistics*, 36(3):1171–1220, 2008. `doi:10.1214/009053607000000677`.

20 Sarang C. Joshi, Raj Varma Kommaraju, Jeff M. Phillips, and Suresh Venkatasubramanian. Comparing distributions and shapes using the kernel distance. In Ferran Hurtado and Marc J. van Kreveld, editors, *Proceedings of the 27th ACM Symposium on Computational Geometry, Paris, France, June 13-15, 2011*, pages 47–56. ACM, 2011.

21 F. Liu, X. Huang, Y. Chen, and J. K. Suykens. Random features for kernel approximation: A survey on algorithms, theory, and beyond. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 44(10):7128–7148, 2022. `doi:10.1109/TPAMI.2021.3097011`.

22 Martin Lotz. Persistent homology for low-complexity models. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475(2230):20190081, 2019. `doi:10.1098/rspa.2019.0081`.

23 K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1-2):1–141, 2017. `doi:10.1561/2200000060`.

24 Cameron Musco and David P. Woodruff. Is input sparsity time possible for kernel low-rank approximation? In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 4438–4448, Red Hook, NY, USA, 2017. Curran Associates Inc.

25 John Nash. C1 isometric imbeddings. *Annals of Mathematics*, 60(3):383–396, 1954.

26 Jeff M. Phillips. ε-samples for kernels. In Sanjeev Khanna, editor, *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 1622–1632. SIAM, 2013.

27 Jeff M. Phillips and Wai Ming Tai. The GaussianSketch for Almost Relative Error Kernel Distance. In Jaroslaw Byrka and Raghu Meka, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2020, August 17-19, 2020, Virtual Conference*, volume 176 of *LIPIcs*, pages 12:1–12:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. `doi:10.4230/LIPIcs.APPROX/RANDOM.2020.12`.

28 Jeff M. Phillips, Bei Wang, and Yan Zheng. Geometric inference on kernel density estimates. In Lars Arge and János Pach, editors, *31st International Symposium on Computational Geometry, SoCG 2015, June 22-25, 2015, Eindhoven, The Netherlands*, volume 34 of *LIPIcs*, pages 857–871. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2015. `doi:10.4230/LIPIcs.SOCG.2015.857`.

29 Rafał Latała. Estimates of Moments and Tails of Gaussian Chaoses. *Ann. Probab.*, 34(6):2315–2331, November 2006.

30 Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In John C. Platt, Daphne Koller, Yoram Singer, and Sam T. Roweis, editors, *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3-6, 2007*, pages 1177–1184. Curran Associates, Inc., 2007. URL: `https://proceedings.neurips.cc/paper/2007/hash/013a006f03dbc5392effeb8f18fda755-Abstract.html`.

31 Donald R. Sheehy. The persistent homology of distance functions under random projection. In Siu-Wing Cheng and Olivier Devillers, editors, *30th Annual Symposium on Computational Geometry, SOCG'14, Kyoto, Japan, June 08 - 11, 2014*, page 328. ACM, 2014. doi:10.1145/2582112.2582126.

32 Bharath K. Sriperumbudur, Kenji Fukumizu, and Gert R. G. Lanckriet. Universality, Characteristic Kernels and RKHS Embedding of Measures. *J. Mach. Learn. Res.*, 12:2389–2410, July 2011.

33 Michel Talagrand. *Gaussian Chaos*, pages 457–492. Springer International Publishing, Cham, 2021. doi:10.1007/978-3-030-82595-9_15.

34 Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi:10.1017/9781108231596.

35 Siddharth Vishwanath, Kenji Fukumizu, Satoshi Kuriki, and Bharath Sriperumbudur. Robust persistence diagrams using reproducing kernels. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc.

36 Shusen Wang, Alex Gittens, and Michael W. Mahoney. Scalable Kernel K-Means Clustering with Nyström Approximation: Relative-Error Bounds. *J. Mach. Learn. Res.*, 20(1):431–479, January 2019.