Inferring ϵ-nets of Finite Sets in a RKHS

Antoine Moniot, Isaure Chauvot de Beauchêne, and Yann Guermeur

LORIA-CNRS, Campus Scientifique, BP 239,
F-54506 Vandœuvre-lès-Nancy Cedex, France,
Antoine.Moniot,Isaure.Chauvot-de-Beauchene,Yann.Guermeur@loria.fr

Abstract. We introduce a method to derive ϵ-nets of finite sets. It operates in a reproducing kernel Hilbert space. Its principle combines two well-known tools of empirical inference: the hierarchical agglomerative clustering and the computation of minimum enclosing balls. It produces ϵ-nets whose cardinalities are smaller than those obtained with state-of-the-art methods.

Keywords: ϵ-nets, empirical inference, RKHS, hierarchical agglomerative clustering

1 Introduction

The inference of sets of prototypes is an ubiquitous problem in applied research. Such sets must satisfy two contradictory constraints: be representative and of cardinality as small as possible. When the data set is included in a Hilbert space (the criterion of fitness is a distance) and its cardinality is finite, then the problem reduces to that of the inference from a Gram matrix of an ϵ-net of minimum cardinality. This theoretical framework is of particular interest, for instance, in structural biology (use of rotamers [1]). Strangely enough, it still calls for the specification of dedicated methods. This is the subject of this study. Without loss of generality, it addresses the case when the dot product is provided by a kernel. Finding an optimal solution, i.e., an ϵ-net whose cardinality is the corresponding covering number, appears as a problem of combinatorial optimization which is intractable in practice. However, feasible solutions can be derived at negligible cost from the dendrograms produced by most of the major algorithms of hierarchical agglomerative clustering (HAC). In that context, a double observation can be made: the corresponding linkage functions are unevenly suited for the task, and it is possible to exhibit a new criterion a priori superior. Our method can be seen as a new HAC algorithm based on the computation of minimum enclosing balls (MEBs). When compared with state-of-the-art HAC algorithms on a classical benchmark, it produces uniformly the smallest ϵ-nets.

The organization of the paper is as follows. The problem is specified in Section 2. Section 3 is devoted to the state of the art. Our method is introduced in Section 4. Section 5 presents the experimental results. At last, we draw conclusions in Section 6.

2 Problem characterization

The problem of interest is a problem of computational geometry. It is formalized as follows.

2.1 Theoretical framework

We first characterize the data available. They constitute a finite subset of a description space $\mathcal{X}$. Let $s_{\mathcal{X}^n} = \{x_i : 1 \leqslant i \leqslant n\}$ denote this subset. A reproducing kernel Hilbert space (RKHS) [2] is spanned from $\mathcal{X}$ by means of a real-valued positive type function/kernel on $\mathcal{X}^2$. Let κ denote this function and $\left(\mathbf{H}_\kappa, \langle\cdot,\cdot\rangle_{\mathbf{H}_\kappa}\right)$ the corresponding RKHS. The canonical distance on $\mathbf{H}_\kappa$, ρ_κ, satisfies :

$$\forall\,(x,x') \in \mathcal{X}^2,\ \rho_\kappa\left(\kappa_x, \kappa_{x'}\right) = \left\|\kappa_x - \kappa_{x'}\right\|_{\mathbf{H}_\kappa} = \sqrt{\kappa\left(x,x\right) + \kappa\left(x',x'\right) - 2\kappa\left(x,x'\right)}.$$

This algebraic framework being established, the knowledge source available for inference can be defined as the Gram matrix $K \in \mathcal{M}_{n,n}\left(\mathbb{R}\right)$ with entries $k_{i,j} = \kappa\left(x_i, x_j\right)$. The hypothesis of a Hilbert space structure and the availability of the matrix K are more than enough to tackle our problem of computational geometry (which becomes as a result a problem of functional analysis). This problem involves concepts whose introduction is usually attributed to Kolmogorov and Tihomirov [3]. Without loss of generality, they are presented below for metric spaces. $(\mathcal{E}, \rho)$ is such a space and $\mathcal{E}'$ is a subset of $\mathcal{E}$ which is totally bounded.

Definition 1 (ϵ-cover, ϵ-net and covering numbers).
For $\epsilon \in \mathbb{R}_+^*$, an ϵ-cover *of $\mathcal{E}'$ is a coverage of $\mathcal{E}'$ with open balls of radius ϵ the centres of which belong to $\mathcal{E}$. These centres form an* ϵ-net *of $\mathcal{E}'$. An* internal/proper ϵ-net *of $\mathcal{E}'$ is an ϵ-net of $\mathcal{E}'$ included in $\mathcal{E}'$. The* covering number *$\mathcal{N}\left(\epsilon, \mathcal{E}', \rho\right)$ is the smallest cardinality of its ϵ-nets:*

$$\mathcal{N}\left(\epsilon, \mathcal{E}', \rho\right) = \min\left\{\left|\mathcal{E}''\right| :\ \left(\mathcal{E}'' \subset \mathcal{E}\right) \wedge \left(\forall e \in \mathcal{E}',\ \rho\left(e, \mathcal{E}''\right) < \epsilon\right)\right\}.$$

$\mathcal{N}^{int}\left(\epsilon, \mathcal{E}', \rho\right)$ will designate a covering number of $\mathcal{E}'$ obtained by considering internal ϵ-nets only. We have thus:

$$\mathcal{N}^{int}\left(\epsilon, \mathcal{E}', \rho\right) = \min\left\{\left|\mathcal{E}''\right| :\ \left(\mathcal{E}'' \subset \mathcal{E}'\right) \wedge \left(\forall e \in \mathcal{E}',\ \rho\left(e, \mathcal{E}''\right) < \epsilon\right)\right\}.$$

The problem considered is the following one.

Problem 1. Given the matrix K and $\epsilon \in \mathbb{R}_+^*$, find an ϵ-net $\mathcal{C}\left(\epsilon\right) = \left\{c_j\left(\epsilon\right) : 1 \leqslant j \leqslant \left|\mathcal{C}\left(\epsilon\right)\right|\right\}$ of $\mathcal{S} = \left\{\kappa_{x_i} : 1 \leqslant i \leqslant n\right\}$ with cardinality $\left|\mathcal{C}\left(\epsilon\right)\right|$ as close as possible to $\mathcal{N}\left(\epsilon, \mathcal{S}, \rho_\kappa\right)$.

2.2 Optimization problem

Problem 1 can be turned into a problem of optimization by means of the following proposition, whose proof is straightforward.

Proposition 1. *The value of $\mathcal{N}(\epsilon, \mathcal{S}, \rho_\kappa)$ is $\mathcal{N}^*$ if and only if $\mathcal{N}^*$ is the minimum natural number such that there exists a partition of $\mathcal{S}$ into $\mathcal{N}^*$ subsets and for each subset, the radius of the MEB is inferior to ϵ.*

In the sequel, $\mathcal{P} = \{p_k : 1 \leqslant k \leqslant |\mathcal{P}|\}$ designates a partition of $\mathcal{S}$ into $|\mathcal{P}|$ subsests p_k. $\mathcal{P}(\mathcal{S})$ designates the set of all these partitions. For every $p_k \in \mathcal{P}$, $c(p_k)$ is the centre of its MEB. Then, $\mathcal{C}(\mathcal{P}) = \{c(p_k) : 1 \leqslant k \leqslant |\mathcal{P}|\}$. With this proposition and notation at hand, an optimal solution to Problem 1, i.e., a solution satisfying $|\mathcal{C}(\epsilon)| = \mathcal{N}(\epsilon, \mathcal{S}, \rho_\kappa)$, is obtained by solving Problem 2.

Problem 2.

$$\min_{\mathcal{P} \in \mathcal{P}(\mathcal{S})} |\mathcal{P}|$$

$$\text{s.t. } \forall k \in [\![1; |\mathcal{P}|]\!],\ \max_{\left\{i : \kappa_{x_i} \in p_k\right\}} \rho_\kappa(\kappa_{x_i}, c(p_k)) < \epsilon.$$

Let $\mathcal{P}^*$ denote an optimal solution to Problem 2. Then according to Proposition 1, $\mathcal{C}(\mathcal{P}^*)$ is an ϵ-net of $\mathcal{S}$ of minimal cardinality ($|\mathcal{P}^*| = \mathcal{N}(\epsilon, \mathcal{S}, \rho_\kappa)$), so that an optimal solution to Problem 1 is obtained by setting $\mathcal{C}(\epsilon) = \mathcal{C}(\mathcal{P}^*)$. Unfortunately, Problem 2 is an intractable problem of combinatorial optimization. To the best our knowledge, the only method available to solve it is the exhaustive way, and the number of partitions of $\mathcal{S}$, i.e., its Bell number B_n, is bounded from below by:

$$B_n \geqslant \left(\frac{n-1}{2}\right)^{\frac{n-1}{4}}.$$

However, the study of Problem 2 remains enlightening to derive a method producing (suboptimal) solutions to Problem 1. Indeed, there is a well-known family of classification methods initially producing as many clusters as there are points (trivial feasible solution to Problem 2) and iteratively merging clusters (producing partitions of smaller and smaller cardinality) according to a dissimilarity measure: the family of HAC algorithms [4]. If one can be satisfied with a suboptimal solution to Problem 2, then these algorithms turn the problem into a tractable one since the number of partitions considered is now upper bounded by n. The question then becomes: how close to $\mathcal{N}(\epsilon, \mathcal{S}, \rho_\kappa)$ is the cardinality of the smallest partition (closest to the root) associated with an ϵ-cover? Of course, the answer depends on the nature of the dissimilarity measure (the *linkage* d_H). The following section discusses the connection between the linkages of the HAC algorithms and the radii of the balls enclosing the sets constituting the partitions.

3 State of the art

In this section, $\mathcal{S}_1$ and $\mathcal{S}_2$ stand for two subsets of $\mathcal{S}$ among the candidates to be merged at some iteration of an HAC algorithm. As such, they are specific subsets of $\mathcal{S}$. We have seen that even though the HAC algorithms are not designed to infer

ϵ-nets, they can all be used for that purpose. It suffices, for each of the nested partitions, to compute the corresponding balls using the matrix K. Furthermore, there exist HAC algorithms which exhibit the following property.

Property 1. The linkage function d_H increases or keeps unchanged each time two clusters are merged and there exists a function f_H such that, for every pair $(\mathcal{S}_1, \mathcal{S}_2)$, $r\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right)$, the radius of the smallest ball enclosing $\mathcal{S}_1 \bigcup \mathcal{S}_2$, is bounded from above by $f_H\left(d_H\left(\mathcal{S}_1, \mathcal{S}_2\right), \left|\mathcal{S}_1\right|, \left|\mathcal{S}_2\right|\right)$.

Property 1 is of central importance since it introduces an explicit connection between the partitions generated and ϵ-nets. Intuitively, the methods which have it are all the more promising to produce small ϵ-nets as $f_H\left(d_H\left(\mathcal{S}_1, \mathcal{S}_2\right), \left|\mathcal{S}_1\right|, \left|\mathcal{S}_2\right|\right)$ is a sharper upper bound on $r\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right)$. Practically, it makes it possible to produce a (non trivial) feasible solution to Problem 2 without computing any ball, i.e., without using K (after the dendrogram has been produced). This amounts to solving Problem 3.

Problem 3. Given an HAC algorithm, the dendrogram it produces for K and $\epsilon \in \mathbb{R}_+^*$, find among the partitions $\mathcal{P}$ for which it can be established that the set $\mathcal{C}\left(\mathcal{P}\right)$ is an ϵ-net of $\mathcal{S}$, the partition closest to the root.

Consequently, when a function f_H is available, a trade-off is to be found between accuracy and computational complexity. Indeed, solving Problem 3 provides us with a (low-cost) initial solution to Problem 2 that can a priori be improved by exploiting K to compute the balls associated with partitions closer to the root. On the other hand, these balls are obtained as the solution of a quadratic programming problem (see Section 4.2).

These observations drive us into studying the main HAC algorithms [5] in the light of Property 1. They are characterized (by means of their linkage) in Table 1.

Table 1. Main HAC algorithms

Method	Linkage d_H
Average	$d_A\left(\mathcal{S}_1, \mathcal{S}_2\right) = \frac{1}{\left\|\mathcal{S}_1\right\|\left\|\mathcal{S}_2\right\|} \sum_{\left(\kappa_x, \kappa_{x'}\right) \in \mathcal{S}_1 \times \mathcal{S}_2} \rho_\kappa\left(\kappa_x, \kappa_{x'}\right)$
Centroid	$d_{Ce}\left(\mathcal{S}_1, \mathcal{S}_2\right) = \rho_\kappa\left(\frac{1}{\left\|\mathcal{S}_1\right\|} \sum_{\kappa_x \in \mathcal{S}_1} \kappa_x, \frac{1}{\left\|\mathcal{S}_2\right\|} \sum_{\kappa_{x'} \in \mathcal{S}_2} \kappa_{x'}\right)$
Complete	$d_{Co}\left(\mathcal{S}_1, \mathcal{S}_2\right) = \max_{\left(\kappa_x, \kappa_{x'}\right) \in \mathcal{S}_1 \times \mathcal{S}_2} \rho_\kappa\left(\kappa_x, \kappa_{x'}\right)$
Minimax	$d_{mM}\left(\mathcal{S}_1, \mathcal{S}_2\right) = \min_{\kappa_x \in \mathcal{S}_1 \bigcup \mathcal{S}_2} \max_{\kappa_{x'} \in \mathcal{S}_1 \bigcup \mathcal{S}_2} \rho_\kappa\left(\kappa_x, \kappa_{x'}\right)$
Single	$d_S\left(\mathcal{S}_1, \mathcal{S}_2\right) = \min_{\left(\kappa_x, \kappa_{x'}\right) \in \mathcal{S}_1 \times \mathcal{S}_2} \rho_\kappa\left(\kappa_x, \kappa_{x'}\right)$

We first note that one single function d_H does not satisfy the monotonicity condition of Property 1: d_{Ce} [6]. We thus restrict the study to the four other

linkages. In the sequel, for each subset $\mathcal{S}'$ of $\mathcal{S}$, $\mathrm{diam}_\kappa\left(\mathcal{S}'\right)$ designates its diameter with respect to the metric ρ_κ. The simplest function f_H to exhibit is obtained for the Minimax linkage. Indeed, since $\mathcal{S}_1 \bigcup \mathcal{S}_2 \subset \mathrm{conv}\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right)$,

$$\begin{aligned}\forall \left(\mathcal{S}_1, \mathcal{S}_2\right),\ r\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right) &= \min_{\kappa_x \in \mathrm{conv}\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right)} \max_{\kappa_{x'} \in \mathcal{S}_1 \bigcup \mathcal{S}_2} \rho_\kappa\left(\kappa_x, \kappa_{x'}\right) \\ &\leqslant d_{mM}\left(\mathcal{S}_1, \mathcal{S}_2\right)\end{aligned}$$

so that we can set:

$$\forall \left(\mathcal{S}_1, \mathcal{S}_2\right),\ f_{mM}\left(d_{mM}\left(\mathcal{S}_1, \mathcal{S}_2\right), \left|\mathcal{S}_1\right|, \left|\mathcal{S}_2\right|\right) = d_{mM}\left(\mathcal{S}_1, \mathcal{S}_2\right).$$

We now turn to the Complete linkage.

$$\begin{aligned}\forall \left(\mathcal{S}_1, \mathcal{S}_2\right),\ r\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right) &\leqslant \frac{1}{\sqrt{2}} \mathrm{diam}_\kappa\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right) \qquad (1) \\ &= \frac{1}{\sqrt{2}} \max_{\left\{\kappa_x, \kappa_{x'}\right\} \in \mathcal{S}_1 \bigcup \mathcal{S}_2} \rho_\kappa\left(\kappa_x, \kappa_{x'}\right) \\ &= \frac{1}{\sqrt{2}} \max_{\left(\kappa_x, \kappa_{x'}\right) \in \mathcal{S}_1 \times \mathcal{S}_2} \rho_\kappa\left(\kappa_x, \kappa_{x'}\right) \qquad (2) \\ &= \frac{1}{\sqrt{2}} d_{Co}\left(\mathcal{S}_1, \mathcal{S}_2\right).\end{aligned}$$

Inequality (1) is an application of Jung's theorem. The transition to (2) (which would be false if $\mathcal{S}_1$ and $\mathcal{S}_2$ were arbitrary subsets of $\mathcal{S}$), springs from the "monotonicity" of d_{Co}. Consequently, we can set

$$\forall \left(\mathcal{S}_1, \mathcal{S}_2\right),\ f_{Co}\left(d_{Co}\left(\mathcal{S}_1, \mathcal{S}_2\right), \left|\mathcal{S}_1\right|, \left|\mathcal{S}_2\right|\right) = \frac{1}{\sqrt{2}} d_{Co}\left(\mathcal{S}_1, \mathcal{S}_2\right).$$

This bound is tight unless we can exploit some knowledge regarding the kernel, for instance the dimensionality of the RKHS. As for the Average linkage, since the corresponding sum can be bounded from below by the maximum, applying exactly the same reasoning as for the Complete linkage produces:

$$\forall \left(\mathcal{S}_1, \mathcal{S}_2\right),\ f_A\left(d_A\left(\mathcal{S}_1, \mathcal{S}_2\right), \left|\mathcal{S}_1\right|, \left|\mathcal{S}_2\right|\right) = \frac{\left|\mathcal{S}_1\right| \left|\mathcal{S}_2\right|}{\sqrt{2}} d_A\left(\mathcal{S}_1, \mathcal{S}_2\right).$$

Regarding the Single linkage, the triangle inequality provides us with

$$\forall \left(\mathcal{S}_1, \mathcal{S}_2\right),\ \mathrm{diam}_\kappa\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right) \leqslant \mathrm{diam}_\kappa\left(\mathcal{S}_1\right) + d_S\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right) + \mathrm{diam}_\kappa\left(\mathcal{S}_2\right).$$

As a consequence, one can prove by induction that

$$\mathrm{diam}_\kappa\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right) \leqslant \left(\left|\mathcal{S}_1 \bigcup \mathcal{S}_2\right| - 1\right) d_S\left(\mathcal{S}_1, \mathcal{S}_2\right).$$

Then, another application of Jung's theorem allows us to define the function f_S as:

$$\forall \left(\mathcal{S}_1, \mathcal{S}_2\right),\ f_S\left(d_S\left(\mathcal{S}_1, \mathcal{S}_2\right), \left|\mathcal{S}_1\right|, \left|\mathcal{S}_2\right|\right) = \frac{\left|\mathcal{S}_1 \bigcup \mathcal{S}_2\right| - 1}{\sqrt{2}} d_S\left(\mathcal{S}_1, \mathcal{S}_2\right).$$

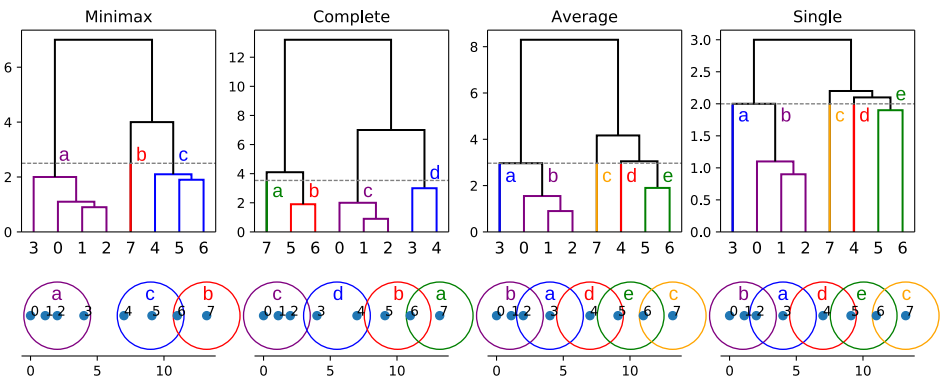


Fig. 1. Dendrograms produced with four linkages: d_{mM}, d_{Co}, d_A and d_S. A cut at the height of the dashed lines provides the corresponding solutions to Problem 3 (for $\epsilon = 2.5$). The resulting ϵ-covers are given below.

Figure 1 illustrates the solutions to Problem 3 obtained with the four linkage functions (d_{mM}, d_{Co}, d_A and d_S) on a set of 8 points of the real line.

The solution of smallest cardinality ($|\mathcal{P}| = 3$) is provided by the Minimax linkage. This solution is optimal since 3 is also the value of the covering number. Using K as an additional knowledge source to cut the dendrograms closer to the root produces new mergings so that eventually, all four ϵ-covers have minimal cardinality. Only one differs from the others: the one obtained with the Complete linkage.

Looking at the functions f_H, the Minimax linkage seems to be the best to produce ϵ-nets of small cardinality. It is noticeable that by definition, its ϵ-nets are internal, which represents an unnecessary restriction that could have a negative effect on the cardinality of the ϵ-net retained (keeping in mind that $\mathcal{N}(\epsilon, \mathcal{E}', \rho) \leqslant \mathcal{N}^{\text{int}}(\epsilon, \mathcal{E}', \rho)$). These observations are at the origin of the specification of our method, which is now introduced.

4 New method of inference

4.1 Definition

Our method for the empirical inference of ϵ-nets is basically the HAC algorithm involving the *radius* linkage function:

$$\forall (\mathcal{S}_1, \mathcal{S}_2), \; d_R(\mathcal{S}_1, \mathcal{S}_2) = r\left(\mathcal{S}_1 \bigcup \mathcal{S}_2\right). \tag{3}$$

The only difference is that the algorithm stops as soon as the candidate merging (minimizing d_R) produces a set whose MEB has a radius superior or equal to ϵ. The prototypes, i.e., the centres of the balls associated with the smallest partition reached, live in a RKHS which can be infinite dimensional (for instance if κ is a radial basis function). However, this raises no difficulties thanks to the kernel trick.

4.2 Derivation of the prototypes

Since the prototypes are the centres of MEBs, each of them is obtained as the solution of an instance of a unique convex quadratic programming (QP) problem. The study of this problem provides us with their analytical expression (in the RKHS). Let $s_{\mathcal{X}^m}$ be a subset of $s_{\mathcal{X}^n}$ of cardinality m. Without loss of generality, its points are reindexed so that we can write $s_{\mathcal{X}^m} = \{x_i : 1 \leqslant i \leqslant m\}$. $K_m \in \mathcal{M}_{m,m}(\mathbb{R})$ is the corresponding Gram matrix and $\boldsymbol{\kappa}_m \in \mathbb{R}_+^m$ is its first diagonal. Then the QP problem is the following one:

Problem 4.

$$\min_{O \in \mathbf{H}_\kappa, R_2 \in \mathbb{R}_+} R_2$$

$$\text{s.t. } \forall i \in [\![1; m]\!],\ \rho_\kappa^2(O, \kappa_{x_i}) \leqslant R_2.$$

Problem 4 can be difficult to solve directly, especially if the RKHS is infinite dimensional. However, the classical way to infer the values of the parameters of a kernel method, the application of the Lagrangian duality, is available. The dual takes the following form:

Problem 5.

$$\min_{\alpha \in \mathbb{R}^m} \left\{\alpha^T K_m \alpha - \boldsymbol{\kappa}_m^T \alpha\right\}$$

$$\text{s.t. } \begin{cases} \forall i \in [\![1; m]\!],\ \alpha_i \geqslant 0 \\ \mathbf{1}_m^T \alpha = 1 \end{cases}.$$

By noting (O^*, R_2^*) the solution to the primal problem and α^* the solution to its dual, we deduce from the Kuhn-Tucker (KT) conditions the analytical expression of the prototype of $s_{\mathcal{X}^m}$:

$$O^* = \sum_{i=1}^m \alpha_i^* \kappa_{x_i}. \tag{4}$$

Precisely, Equation (4) provides the location of the centre of the MEB of $s_{\mathcal{X}^m}$ in its convex hull. With this formula at hand, the square distance between a prototype and any point in $\mathbf{H}_\kappa$ (let it be in $\mathcal{S}$ or not) is

$$\begin{aligned} \rho_\kappa^2(O^*, \kappa_x) = &\ \|O^*\|_{\mathbf{H}_\kappa}^2 + \|\kappa_x\|_{\mathbf{H}_\kappa}^2 - 2\langle O^*, \kappa_x \rangle_{\mathbf{H}_\kappa} \\ = &\ \sum_{i=1}^m \sum_{j=1}^m \alpha_i^* \alpha_j^* \kappa(x_i, x_j) + \kappa(x, x) - 2\sum_{i=1}^m \alpha_i^* \kappa(x_i, x). \end{aligned}$$

As for the value of the radius $\sqrt{R_2^*}$, since $\mathbf{1}_m^T \alpha^* = 1$, the KT complementary conditions provide us with:

$$R_2^* = \sum_{i=1}^m \alpha_i^* \rho_\kappa^2(O^*, \kappa_{x_i}).$$

4.3 Algorithmic implementation

The literature provides us with several exact algorithms to compute an MEB (see [7] for a survey). Their time complexity is a $O\left(m^3\right)$. However, they only apply to data sets living in a finite dimensional space (they solve Problem 4). This is precisely to bypass this limitation that we solve the dual (Problem 5) instead. It is the minimization of a quadratic form over the standard/probability simplex. This property can be exploited by a classical QP method, the Frank-Wolfe algorithm [8], which happens to be among the state-of-the-art "approximation algorithms" for the computation of MEBs. At each iteration, the solution of the linear programming problem is the vertex of the standard simplex whose index is the one of the smallest component of the gradient. Furthermore, there is an analytic expression for the optimal step size. Our implementation of the method makes use of both features. In practice, for a value of the duality gap equal to 0.99, the time complexity observed is a $O\left(m^2\right)$.

5 Experimental results

Our method is evaluated in the framework of a comparative study. The criterion of assessment is the cardinality of the ϵ-nets generated. The methods of reference are the four HAC algorithms of Section 3 which exhibit Property 1. Thus, the study compares the efficacy (to produce small ϵ-nets) of the following linkage functions: d_R (given by Equation 3), d_A, d_{Co}, d_{mM} and d_S (gathered in Table 1). Our method is implemented as described in Section 4. As for the algorithms from the literature, the ϵ-nets are obtained by solving Problem 3 based on the values returned by the corresponding functions f_H.

The benchmark selected is one of those used by Bien and Tibshirani in [5]: the Olivetti Faces data set[1]. It is made up of images of the faces of 40 different individuals. For each individual, 10 images are provided. They are 8-bit grayscale images of size 64×64. Thus, the data are vectors of $\mathbb{R}^{64 \cdot 64}$. This space is endowed with its canonical structure of Euclidean space. To sum up, the algorithms operate on the Gram matrix of $s_{\mathcal{X}^n} = \{x_i : 1 \leqslant i \leqslant n\}$ where $n = 400$. It is possible to infer from this matrix noteworthy elements: the two closest images, the two most distant images, the center of the MEB, and the images on its surface. Those elements are represented in Fig 2.

Let $d_{2,\min} = \min_{1 \leqslant i < j \leqslant n} d_2(x_i, x_j)$. Then, $d_{2,\min}$, $r\left(s_{\mathcal{X}^n}\right)$, and $\operatorname{diam}\left(s_{\mathcal{X}^n}\right)$ can be used to restrict the study to the values of ϵ for which the derivation of an ϵ-net is non trivial. Indeed, if $\epsilon < \frac{1}{2} d_{2,\min}$, then $\mathcal{N}\left(\epsilon, s_{\mathcal{X}^n}, d_2\right) = n$ and we can set $\mathcal{C}\left(\epsilon\right) = s_{\mathcal{X}^n}$. Conversely, if $\epsilon > r\left(s_{\mathcal{X}^n}\right)$, then $\mathcal{N}\left(\epsilon, s_{\mathcal{X}^n}, d_2\right) = 1$ and we can set $\mathcal{C}\left(\epsilon\right) = c\left(s_{\mathcal{X}^n}\right)$ (centre of the MEB of $s_{\mathcal{X}^n}$). At last, if $\epsilon > \operatorname{diam}\left(s_{\mathcal{X}^n}\right)$, then $\mathcal{N}^{int}\left(\epsilon, s_{\mathcal{X}^n}, d_2\right) = 1$ and we can set $\mathcal{C}\left(\epsilon\right)$ equal to any singleton.

Consequently, the cardinalities of the ϵ-nets produced by the five algorithms are computed for ϵ in the interval $(0, \operatorname{diam}\left(s_{\mathcal{X}^n}\right)]$. The corresponding curves are gathered in Fig 3.

[1] https://cs.nyu.edu/~roweis/data.html

Fig. 2. Olivetti Faces dataset. (a) and (b): two closest images; (c) and (d): two most distant images; (e): center of the MEB; (f): one image on the surface of the MEB.

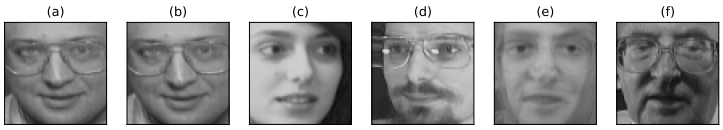
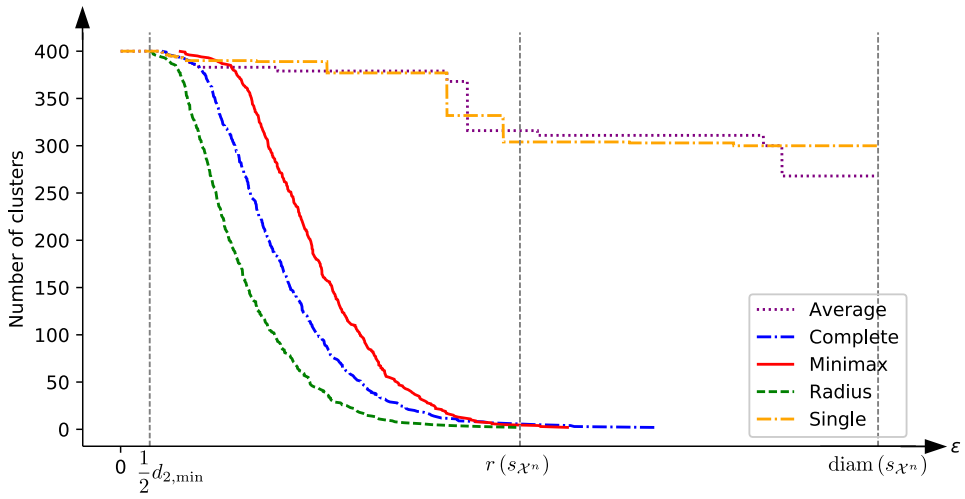


Fig. 3. Cardinality of the ϵ-nets produced by HAC algorithms as a function of ϵ



The new method is uniformly (for all values of ϵ) superior to the other four. Focusing on the HAC algorithms from the literature, it is noteworthy that contrary to our intuition, the Complete linkage appears to be more efficient than the Minimax linkage (at least for small values of ϵ). This phenomenon is the subject of ongoing investigations. As expected, neither d_A nor d_S appears as an appropriate choice.

6 Conclusions

A method to infer prototypes/ϵ-nets for finite subsets of a Hilbert space has been introduced. It can be seen as a hierarchical agglomerative clustering algorithm that operates in a reproducing kernel Hilbert space. Its linkage function takes in input two clusters of a partition and returns the radius of the minimum ball enclosing their union. The prototypes inferred can be easily handled in the RKHS, even when this space is infinite dimensional. A comparative experimental study involving the most challenging HAC algorithms from the literature suggests that our method produces smaller ϵ-nets, i.e., behaves as required.

Our current work is two-fold. It consists in performing additional comparative experiments and performing kernel engineering so as to dedicate the method to different applications in structural biology [9].

References

1. Murray, L.J.W., Arendall, W.B., Richardson, D.C., Richardson, J.S.: RNA backbone is rotameric. Proceedings of the National Academy of Sciences **100**(24), 13,904–13,909 (2003)
2. Berlinet, A., Thomas-Agnan, C.: Reproducing Kernel Hilbert Spaces in Probability and Statistics. Kluwer Academic Publishers, Boston (2004)
3. Kolmogorov, A., Tihomirov, V.: ϵ-entropy and ϵ-capacity of sets in functional spaces. American Mathematical Society Translations, series 2 **17**, 277–364 (1961)
4. Gower, J.C., Ross, G.J.S.: Minimum spanning trees and single linkage cluster analysis. Journal of The Royal Statistical Society Series C-applied Statistics **18**, 54–64 (1969)
5. Bien, J., Tibshirani, R.: Hierarchical clustering with prototypes via minimax linkage. Journal of the American Statistical Association **106**(495), 1075–1084 (2011)
6. Kopp, B.: Hierarchical classification III: Average-linkage, median, centroid, WARD, flexible strategy. Biometrical Journal **20**, 703–711 (1978)
7. Källberg, L.: Minimum enclosing balls and ellipsoids in general dimensions. Ph.D. thesis, Mälardalen University (2019)
8. Frank, M., Wolfe, P.: An algorithm for quadratic programming. Naval Research Logistics Quarterly **3**(1-2), 95–110 (1956)
9. Delannoy, A., Moniot, A., Guermeur, Y., Chauvot de Beauchêne, I.: Feature extraction for the clustering of small 3D structures: application to RNA fragments. In: JOBIM'21 (2021)