

---

# Active Nearest Neighbor Regression Through Delaunay Refinement

---

Alexander Kravberg<sup>\*1</sup> Giovanni Luca Marchetti<sup>\*1</sup> Vladislav Polianskii<sup>\*1</sup> Anastasiia Varava<sup>2</sup>  
Florian T. Pokorny<sup>1</sup> Danica Kragic<sup>1</sup>

## Abstract

We introduce an algorithm for active function approximation based on nearest neighbor regression. Our Active Nearest Neighbor Regressor (ANNR) relies on the Voronoi-Delaunay framework from computational geometry to subdivide the space into cells with constant estimated function value and select novel query points in a way that takes the geometry of the function graph into account. We consider the recent state-of-the-art active function approximator called DEFER, which is based on incremental rectangular partitioning of the space, as the main baseline. The ANNR addresses a number of limitations that arise from the space subdivision strategy used in DEFER. We provide a computationally efficient implementation of our method, as well as theoretical halting guarantees. Empirical results show that ANNR outperforms the baseline for both closed-form functions and real-world examples, such as gravitational wave parameter inference and exploration of the latent space of a generative model.

## 1. Introduction

The need to *actively* approximate a function by iteratively querying novel points from its domain appears in a variety of theoretical and experimental areas, such as modern physics and astronomy (Ashton et al., 2019; Thrane & Talbot, 2019), chemistry (Gubaev et al., 2018; Ramakrishnan et al., 2015) and density estimation (Sung & Niyogi, 1995; Järvenpää et al., 2019; Kandasamy et al., 2015). Typically, the function to be approximated is not explicitly known, but can be evaluated at arbitrary points from its domain. Such formulation poses two intertwined problems: selecting points for

<sup>\*</sup>Equal contribution <sup>1</sup>School of Electrical Engineering and Computer Science, Royal Institute of Technology (KTH), Stockholm, Sweden. <sup>2</sup>No affiliation. Correspondence to: Giovanni Luca Marchetti <glma@kth.se>.

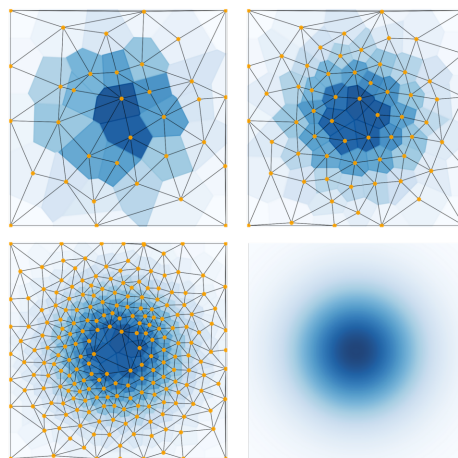


Figure 1: The ANNR progressively approximating a Gaussian function. The approximation is depicted in blue-scale, with the ground-truth function plotted in the lower-right corner. The Delaunay triangulation is depicted in black as well.

evaluation (queries), and performing interpolation given a dataset with known function values. For the former, naive approaches such as sampling uniformly from the domain are computationally infeasible, especially for sparse and high-dimensional functions. Function evaluation is often expensive in real-life applications, as querying a single data point might require running a time- or resource-consuming experiment. It is thus crucial to design efficient strategies for selecting points to evaluate the function upon.

Recently, a scalable function approximator employing active querying has been proposed under the name DEFER (Bodin et al., 2021). It relies on a rectangular partitioning of the ambient space with the datapoints being the centers of the partitions, and approximates the (density) function piecewise constantly on each rectangle. DEFER outperforms state-of-the-art sampling methods in parameter inference tasks as well as in arbitrary function approximation. It is worth noting that despite Bodin et al. (2021) generally address density estimation problems, DEFER is effectively an active function approximator and differs from traditional density estimators (Altman, 1992) that are designed for

static data. Using a rectangular partitioning, however, has a number of disadvantages. Rectangular approximations are not optimal for arbitrary shapes, especially in high dimensions. Indeed, shape approximation via rectangles becomes progressively worse as dimensions grow, which can be illustrated in a well-known ‘spiky rectangle phenomenon’ (Devi et al., 2016).

We instead propose to upgrade the *Nearest Neighbour Regressor* (NNR) (Altman, 1992) to an active setting. The NNR is a function approximator which is locally constant on the Voronoi tessellation. The space is thus partitioned into Voronoi cells, which are arbitrary polytopes adaptive to the local geometry of data (Fortune, 1995). Such a space partitioning and the corresponding locally-constant approximation address the aforementioned disadvantages of DEFER, which we empirically show in the present work.

The core idea behind our active querying strategy is to look for points where the estimated function presents the largest variation. Such points are the most informative for the update of the approximator. This is done by considering the Delaunay triangulation (Fortune, 1995), which is dual to the Voronoi tessellation. The triangulation allows to discretize the graph of the function and to look at its volume, which captures the function variation. Working with the graph of the function allows to balance between the exploration and exploitation strategies. This leads to a geometry-aware procedure deemed *Active Nearest Neighbour Regressor* (ANNR). Voronoi tessellations and their dual Delaunay triangulations enable to solve both the aforementioned problems of interpolation and querying within a single geometric framework.

To make the computation of the Delaunay triangulation feasible in high dimensions, we build upon on the approximate stochastic method described in (Polianskii & Pokorny, 2020). Additionally, we prove via geometric arguments that the volumes of the Delaunay simplices over the graph of  $f$  get arbitrary small as the procedure progresses, obtaining halting guarantees for the ANNR. Our implementation of the ANNR is available at <https://github.com/vlpolyansky/annr>.

Our main contributions can be summarized as follows: (i) a novel active querying procedure deemed ANNR exploiting the geometry of the graph of  $f$  through the Delaunay triangulation; (ii) an efficient high-dimensional implementation and a theoretical proof of halting for the ANNR; (iii) an empirical investigation of the ANNR through a series of experiments demonstrating improved performance and robustness over the recently introduced active function approximator DEFER.

## 2. Related Work

**Delaunay Triangulations.** The Delaunay triangulations were originally introduced in Delaunay et al. (1934) as natural triangulations of point-clouds in arbitrary dimension. Because of their remarkable geometrical properties, they have seen extensive applications within computer graphics (Bern & Eppstein, 1992; Chen, 2004) and topological data analysis (Edelsbrunner & Harer, 2010). The ANNR is in particular related to Delaunay-based techniques for mesh refinement (Shewchuk, 2002), whose goal is to refine the Delaunay triangulation (mesh) of a coarse point-cloud by progressively adding novel points. Ruppert’s algorithm (Ruppert, 1995) and Chew’s second algorithm (Chew, 1993) are the most popular to this end. The idea underlying both algorithms is to insert the circumcenters of poor quality Delaunay triangles. Although our method follows a similar pattern, we are concerned with the task of function approximation rather than mesh refinement. We thus consider the known values of the ground-truth function and query points in order to refine the approximation, rather than aiming to optimally fill the ambient space. Moreover, mesh refinement algorithms and applications to computer graphics in general are concerned with a two- or three-dimensional ambient space. In contrast, we provide a general framework and an efficient implementation of the ANNR in high dimensions.

**Deep Active Learning.** Active querying strategies have been extensively studied in the context of deep learning. However, the vast majority of methods assume a predefined finite pool from which points can be queried (‘pool-based active learning’) or that datapoints can be sampled from the ground-truth distribution and then rejected for querying (‘stream-based active learning’) (Kontorovich et al., 2016; Ren et al., 2021; Settles, 2009). These methods typically deploy an ‘acquisition function’ representing some form of uncertainty that the desired query has to maximize (Lewis & Gale, 1994; MacKay, 1992; Cohn et al., 1994). Picking an arbitrary point in the ambient space –sometimes referred to as ‘membership query synthesis’– is rarely considered in the literature and the corresponding subfield of active learning is relatively undeveloped. The reason is twofold: first, maximizing an acquisition function in the continuum for black-box models such as deep neural networks would require an additional expensive optimization procedure (gradient descent, for example) at each querying step. In contrast, simple but interpretable function approximators such as the piecewise linear ones considered in this work and in DEFER (Bodin et al., 2021) allow to design querying strategies without the need of an acquisition function. Second, an arbitrary datapoint in the ambient space is likely to result in noise and thus to be uninformative to query, if even possible (Baum & Lang, 1992). Due to recent advances in generative modeling, this can be nowadays amended by looking for points to query in the latent space of a generative model

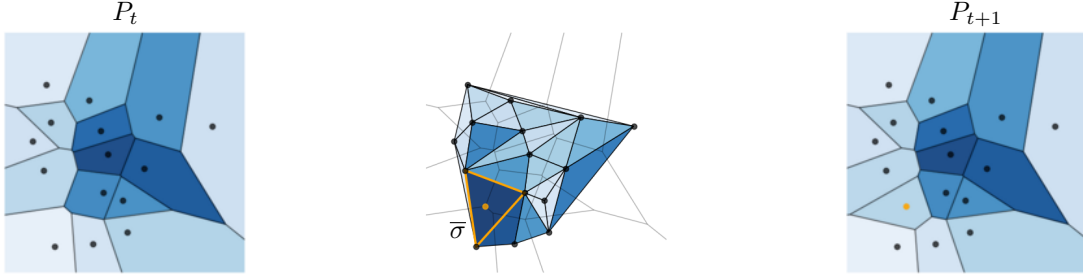


Figure 2: A graphical depiction of the ANNR querying procedure. **Left:** the NNR (in blue-scale) of a two-dimensional dataset. **Center:** the corresponding Delaunay triangulation. The simplices are colored by the volume of their liftings, with the largest one highlighted in orange together with its circumcenter. **Right:** the updated NNR after the orange point has been added to the dataset.

(Zhu & Bento, 2017).

**Active Sampling in Bayesian Inference.** Acquisition functions based on uncertainty naturally occur in Bayesian inference, which is often used to approximate unknown distributions (Sung & Niyogi, 1995). Classical Bayesian methods have a number of disadvantages, such as assuming a certain form of a distribution, requiring computable gradients, and intractability. Recent developments in approximate Bayesian computation and Monte Carlo sampling methods address some of these issues (Järvenpää et al., 2020; Acerbi, 2018; Higson et al., 2019), albeit suffering from impractical computational complexity (Liu et al., 2020). When compared to DEFER, state-of-the-art Bayesian sampling methods showed comparable or worse performance, in particular, on multi-modal, sparse and discontinuous distributions (Bodin et al., 2021). In addition, these methods can only be applied to a particular class of likelihood functions, and are typically designed to perform sampling from the estimated distributions rather than to estimate a (likelihood) function value at an arbitrary point.

### 3. Method

Let  $f : \mathcal{X} \rightarrow \mathbb{R}$  be a function defined on a connected metric space  $(\mathcal{X}, d)$ , where  $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$  denotes the distance. Given a finite set of points  $P \subseteq \mathcal{X}$  (referred to as datapoints) on which the values of  $f$  are known, a natural way to extend  $f$  beyond  $P$  is to regress to the values at the nearest datapoint (Altman, 1992).

**Definition 3.1.** Let  $x$  be a point such that all the distances  $d(x, p)$ ,  $p \in P$  are distinct. The value of the *Nearest Neighbor Regressor* (NNR)  $\tilde{f}$  at  $x$  coincides with the value of  $f$  at the closest datapoint i.e.,

$$\tilde{f}(x) = f(\bar{p}), \quad \bar{p} = \operatorname{argmin}_{p \in P} d(x, p). \quad (1)$$

Recall that the *Voronoi cell*  $C(p)$  of  $p \in P$  contains the points in  $\mathcal{X}$  that are closer to  $p$  than to any other datapoint

i.e.,

$$C(p) = \{x \in \mathcal{X} \mid \forall q \in P \quad d(x, q) \geq d(x, p)\}. \quad (2)$$

The Voronoi cells are closed, cover the ambient space  $\mathcal{X}$  and intersect at their boundary. When  $\mathcal{X} = \mathbb{R}^m$ , they are arbitrary  $m$ -dimensional convex polytopes (Fortune, 1995). From the point of view of the Voronoi tessellation, the NNR is defined on the interior of Voronoi cells and is constant locally therein.

#### 3.1. Active Nearest Neighbor Regression

In this work we upgrade the NNR to an active setting. A general active procedure consists in updating a dataset inductively by querying for the value of  $f$  at a new datapoint based on the current dataset, on which  $f$  is assumed to be known. Starting from an initial dataset  $P_0$ , this produces a sequence  $\{P_t\}_{t \in \mathbb{N}}$  obtained by adding a datapoint at each step:  $P_{t+1} = P_t \cup \{p_t\}$ . We are now going to describe our proposed querying strategy.

To update the dataset, we first consider the triangulation which is dual to the Voronoi tessellation. From now on, we focus on the  $m$ -dimensional Euclidean space  $\mathcal{X} = \mathbb{R}^m$ . In the following we denote by  $\langle \cdot \rangle$  the convex hull of a set.

**Definition 3.2.** The *Delaunay triangulation*  $\operatorname{Del}_P$  generated by  $P$  is the simplicial complex with vertices in  $P$  that contains a  $k$ -dimensional simplex  $\sigma = \langle v_0, \dots, v_k \rangle$ ,  $v_i \in P$ , if and only if

$$\bigcap_{0 \leq i \leq k} C(v_i) \neq \emptyset. \quad (3)$$

If  $P$  is in general position then the simplices in  $\operatorname{Del}_P$  are non-degenerate. In that case, a remarkable property of the Delaunay triangulation is that no point in  $P$  lies inside the circumsphere of an  $m$ -dimensional simplex  $\sigma \in \operatorname{Del}_P$  (Fortune, 1995).

Our querying strategy intuitively relies on the variation of  $f$

over the  $m$ -dimensional simplices of the Delaunay triangulation. Concretely, we compute the volume of the 'lifting' of such simplices to the graph  $\Gamma_{\lambda f} = \{(x, \lambda f(x)) \mid x \in \mathcal{X}\}$  of  $\lambda f$ , where  $\lambda \in \mathbb{R}_{\geq 0}$  is a hyperparameter.

**Definition 3.3.** The *lifting* via  $\lambda f$  of an  $m$ -dimensional simplex  $\sigma = \langle v_0, \dots, v_m \rangle$  in  $\mathbb{R}^m$  is the simplex in  $\mathbb{R}^{m+1}$

$$\hat{\sigma} = \langle (v_0, \lambda f(v_0)), \dots, (v_m, \lambda f(v_m)) \rangle. \quad (4)$$

Our algorithm looks for the  $m$ -dimensional simplex  $\sigma \in \text{Del}_P$  that maximizes  $\text{Vol}(\hat{\sigma})$ . To this end, the  $k$ -dimensional volume of a simplex  $\sigma = \langle v_0, \dots, v_k \rangle$  can be efficiently computed even for high dimensions via the *Cayley-Menger determinant* (Sommerville, 1958):

$$\text{Vol}(\sigma) = \sqrt{\frac{(-1)^{k+1}}{2^k (k!)^2} \det M}. \quad (5)$$

Here,  $M$  is the  $(k+2) \times (k+2)$  matrix obtained by padding by a top row and a left column equal to  $(0, 1, \dots, 1)$  the matrix of mutual distances  $d(v_i, v_j)^2$ .

The role of the hyperparameter  $\lambda$  is to 'sharpen' the lifted simplices. Since  $\lambda$  controls the increment of  $\lambda f$ ,  $\text{Vol}(\hat{\sigma})$  depends monotonically on  $\lambda$ . As  $\text{Vol}(\sigma)$  is constant,  $\lambda \gg 0$  encourages simplices with high variation even if the base simplex  $\sigma$  is small. On the other hand,  $\text{Vol}(\sigma) \sim \text{Vol}(\hat{\sigma})$  for  $\lambda \sim 0$ , in which case the ANNR regularly explores the domain by looking for the largest simplices, disregarding  $f$  and its variation completely. In summary,  $\lambda$  can be interpreted as governing a trade-off between exploitation of the estimated variation and domain exploration – a classical compromise in active learning (Bondu et al., 2010). We make the latter statement formal in Sec. 4.1 and refer to Sec. 5.1 for a further discussion around the practical choice of  $\lambda$ .

Given the simplex  $\bar{\sigma}$  that maximizes  $\text{Vol}(\hat{\sigma})$ , the point we query is the dual of  $\bar{\sigma}$  in the Voronoi tessellation. In other words, the novel query is the value of  $f$  at the intersection between the  $m+1$  Voronoi cells of the vertices of  $\bar{\sigma}$ . It is thus a point where the NNR (Eq. 1) is maximally discontinuous, which motivates the need to gain information on  $f$  around it. Geometrically, it coincides with the *circumcenter* of  $\bar{\sigma}$  (i.e., the point in  $\mathbb{R}^m$  equidistant from the vertices of  $\bar{\sigma}$ ) and we consequently denote it by  $\text{Circ}(\bar{\sigma})$ . Our querying strategy is graphical depicted in Fig. 2 and can be formally summarized as follows:

$$p_{t+1} = \text{Circ}(\bar{\sigma}), \quad \bar{\sigma} = \operatorname{argmax}_{\sigma \in \text{Del}_{P_t}} \text{Vol}(\hat{\sigma}). \quad (6)$$

We stop the iteration when the maximum volume of a lifting reaches a given threshold  $\varepsilon > 0$ . If care is taken in order to

---

**Algorithm 1** Active Nearest Neighbor Regression (ANNR)
 

---

```

Initialize  $P$ 
 $maxVol = \varepsilon$ 
while  $maxVol \geq \varepsilon$  do
    Compute the Delaunay triangulation  $\text{Del}_P$ 
     $maxVol = 0$ 
    for  $\sigma \in \text{Del}_P$  of dimension  $m$  do
        Compute the volume  $\text{Vol}(\hat{\sigma})$  via Eq. 5
        if  $\text{Vol}(\hat{\sigma}) > maxVol$  then
             $maxVol = \text{Vol}(\hat{\sigma})$ 
             $maxSimplex = \hat{\sigma}$ 
        end if
    end for
    Add the circumcenter of  $maxSimplex$  to  $P$ 
end while
    
```

---

constrain the dataset in a compact region (which is discussed in Sec. 3.2), the algorithm is guaranteed to halt with mild assumptions on  $f$  (see Sec. 4.2). The overall procedure is summarized in the pseudocode Alg. 1.

### 3.2. Bounding the Query

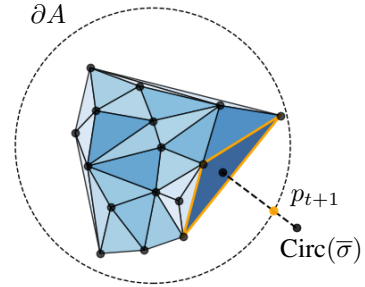


Figure 3: Depiction of the query in the case the circumcenter falls outside of the bounding region  $A$ .

The circumcenter of a simplex can possibly lie outside of the simplex itself. When a  $m$ -simplex is close to being degenerate (i.e., contained in a  $(m-1)$ -dimensional affine subspace), the circumcenter tends to escape in an infinite direction. It is thus necessary to limit the expansion of the dataset within a compact region.

We propose to fix a convex compact set  $A \subseteq \mathbb{R}^m$  containing the initialization  $P_0$  and to constrain  $P_t$  to be contained in  $A$  in the following way. Suppose that  $\bar{\sigma}$  is the simplex in  $\text{Del}_{P_t}$  whose lifting has the largest volume and that  $\text{Circ}(\bar{\sigma}) \notin A$ . Instead of setting  $p_{t+1} = \text{Circ}(\bar{\sigma})$  we set  $p_{t+1}$  as the intersection between the boundary  $\partial A$  of  $A$  and the segment connecting the barycenter of  $\bar{\sigma}$  (which is contained in  $\bar{\sigma}$ ) and  $\text{Circ}(\bar{\sigma})$ . Such segment is known as 'Euler line' of  $\bar{\sigma}$  (Kimberling, 1998) and the intersection between it and  $\partial A$

is unique due to the convexity of  $A$ . The intuition behind this choice for  $p_{t+1}$  is that it is the furthest point in  $A$  from (the barycenter of)  $\bar{\sigma}$  in the direction of the circumcenter of the latter. Moreover, the theoretical halting guarantees discussed in Sec. 4.2 crucially rely on the procedure described here. A graphical depiction is provided in Fig. 3.

The bounding set  $A$  additionally enables to initialize the dataset  $P_0$  by uniformly sampling from  $A$ . Since the Delaunay triangulation covers the convex hull of its vertices, one can additionally enlarge  $P_0$  so that  $\langle P_0 \rangle = A$  and thus  $\langle P_t \rangle = A$  for every  $t$ . This is always possible if  $A$  is polytopal. When  $A$  is a hypercube (which is the case in all of our experiments) this can be done by adding its  $2^m$  vertices to  $P_0$ . We stick to this form of initialization when implementing Alg. 1.

### 3.3. Computing the Delaunay Triangulation

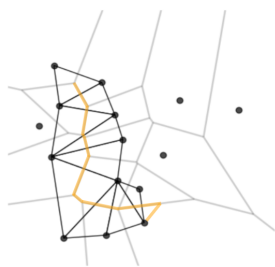


Figure 4: A depiction of a random walk along the boundary of the Voronoi cells (orange) together with the Delaunay simplices found along the way.

A naive implementation of ANNR involving standard exact methods of constructing a Delaunay triangulation would present a significant computational challenge. Namely, the Delaunay triangulation is prohibitively expensive to compute in high dimensions as the number of simplices grows exponentially with respect to the dimension  $m$  in a general scenario (Klee, 1979).

We propose to approximately infer  $\text{Del}_P$  via an adaptation of the stochastic ray-casting technique first considered in Rushdi et al. (2017) and later expanded in Polianskii & Pokorný (2020). The central idea lies in performing a random walk along the boundary of the Voronoi cells in order to look for their vertices, which in turn correspond by duality to the desired Delaunay simplices. More precisely, a Markov Chain (MC) is constructed in the following way. First, a starting datapoint  $x_0$  is picked uniformly from  $P$ . Then a random direction  $\theta_0 \in \mathbb{S}^{m-1}$  is chosen and the intersection  $x_1$  between the ray  $\{x_0 + t\theta_0\}_{t \in \mathbb{R}_{\geq 0}}$  cast from  $x_0$  and a  $(m-1)$ -dimensional face of the polytopal Voronoi cell  $C(x_0)$  is found by an explicit analytic expression. The procedure is repeated by additionally constraining

the subsequent ray to lie on the aforementioned face, obtaining a point  $x_2$  on a  $(m-2)$ -dimensional face of  $C(x_0)$ , and so on. After  $m$  rounds of iteration, a vertex  $x_m$  of the Voronoi cell (i.e., a 0-dimensional face) is obtained. By keeping track of the other Voronoi cells to which the encountered faces belong to, the (vertices of the) Delaunay simplex corresponding to  $x_m$  are automatically available. After that, the random walk continues on the 1-skeleton of the Voronoi tessellation in an analogous manner, possibly departing from  $C(x_0)$  and finding other Delaunay simplices. The result is a subset of the Delaunay triangulation which approximates the whole  $\text{Del}_P$ . By deploying a nearest-neighbor lookup structure such as a  $k$ -d tree (Bentley, 1975), every ray intersection described above can be performed with a complexity that depends *logarithmically* on the number of datapoints. We refer to Polianskii & Pokorný (2020) for further details. A graphical depiction of the described procedure is presented in Fig. 4.

The vertices of the Voronoi cells corresponding to the found Delaunay simplices (i.e., the circumcenters of the latter) are obtained as a byproduct, thus there is no need for a further computation of the point to query. Because of the iterative nature of the search for Delaunay simplices, we integrate the latter with the main loop in Alg. 1 for a further computational improvement. We further heuristically adjust the initialization of the random walks by picking datapoints close to barycenters of the simplices with the highest lifted volume at the previous step. As suggested in (Polianskii & Pokorný, 2020), this can be accomplished via a ‘visibility walk’ (Devillers et al., 2001) in the new 1-skeleton towards such barycenters before initiating the random walk.

### 3.4. Complexity Analysis

In this section we provide a comparison of the computational complexity of the ANNR and DEFER for both data querying and evaluation of the approximated function.

Since each random walk has logarithmic complexity (Sec. 3.3), querying a new datapoint has complexity  $\mathcal{O}(L \log |P|)$  for the ANNR, where  $L$  is the number of MC steps performed. The latter is a hyperparameter typical for MC methods and its effect can be mitigated by running multiple walks in parallel. For DEFER, querying has complexity  $\mathcal{O}(\log |P|)$ . The difference in complexity due to  $L$  indeed reflects the price of approximating the Delaunay triangulation, as opposed to the exact geometry of rectangular partitions in DEFER. Additionally, the ANNR computes volumes of simplices, which has complexity  $\mathcal{O}(m^3)$  w.r.t. the dimension  $m$  due to the Cayley-Menger determinant (Eq. 5), in contrast to linear complexity for volumes of rectangles in DEFER. Overall, querying complexity of the ANNR is  $\mathcal{O}(L(m^3 + \log |P|))$ .

Evaluating the approximated function has identical complexity  $\mathcal{O}(\log |P|)$  with respect to the current dataset size  $|P|$  in both methods. This is due to data structures such as  $k$ -d trees underlying both the nearest neighbor lookup for Voronoi cells in the ANNR and the rectangle lookup in DEFER.

## 4. Theoretical Results

### 4.1. Geometric Interpretation

In this section we present a geometric interpretation of the ANNR based on an inequality for volumes of graphs of functions. To this end, suppose that  $\Omega \subseteq \mathbb{R}^m$  is an  $m$ -dimensional connected and compact submanifold (with boundary). For a smooth function  $f : \Omega \rightarrow \mathbb{R}$  denote by  $\Gamma_f$  its graph, which is an  $m$ -dimensional manifold. Additionally, denote by  $f_\Omega = \frac{1}{\text{Vol}(\Omega)} \int_\Omega f$  the average of  $f$  over  $\Omega$  and by  $g_f$  the matrix  $\nabla_f \otimes \nabla_f$  i.e.,  $(g_f)_{i,j} = \partial_{x_i} f \partial_{x_j} f$ .

**Proposition 4.1.** *There exists a constant  $C > 0$  such that for every smooth function  $f : \Omega \rightarrow \mathbb{R}$  the following inequality holds:*

$$\log \text{Vol}(\Gamma_f) \geq C \|f - f_\Omega\|_2^2 + \log \text{Vol}(\Omega) + o(\|g_f\|^2). \quad (7)$$

We refer to the Appendix for a proof. If  $\Omega$  is a Voronoi cell then the NNR takes the value of  $f$  at the only datapoint contained in the cell, which is a one-sample Monte Carlo estimate of  $f_\Omega$ . In other words,  $f_\Omega \sim \tilde{f}$  on  $\Omega$ . In light of Proposition 4.1, the (logarithm of the) volume of the graph of  $f$  is approximately bounding the error between  $f$  and its NNR estimation plus a term  $\log \text{Vol}(\Omega)$  penalizing domains with large volume. The latter can be thought as a form of regularization. The ANNR exploits this as it computes and approximation to the volume of the graph of (a multiple of)  $f$ . Such volume is not directly approximable when  $\Omega$  is a Voronoi cell because of lack of data in  $\Omega$ . We thus compute it over the discretization of the manifold  $\Gamma_f$  given by the lifted Delaunay triangulation  $\{\hat{\sigma}\}_{\sigma \in \text{Del}_P}$ . In other words, the volume is computed for the simplices formed by the datapoints in neighboring intersecting cells.

Note that if we replace  $f$  by  $\lambda f$  for some  $\lambda \in \mathbb{R}_{\geq 0}$  then the right hand-side of Eq. 7 becomes  $C\lambda^2 \|f - f_\Omega\|_2^2 + \log \text{Vol}(\Omega) + \lambda^4 o(\|g_f\|^2)$ . A natural interpretation is that  $\lambda$  controls the balance between error minimization (corresponding to the term  $\|f - f_\Omega\|_2^2$ ) and exploration of the domain (corresponding to the regularization term  $\log \text{Vol}(\Omega)$ ). This motivates the insertion of the hyperparameter  $\lambda$  in the ANNR (see Eq. 4).

### 4.2. Halting Guarantees

In this section we establish formal halting guarantees for the ANNR given a continuity assumption on the ground-

truth function  $f$ . Since our algorithm stops when a fixed threshold  $\varepsilon$  is met, halting can be equivalently reformulated as the vanishing of the corresponding lower limit. The halting guarantee is then given by the following.

**Proposition 4.2.** *Assume that the ground-truth function  $f$  is Lipschitz and let  $s_t = \max_{\sigma \in \text{Del}_{P_t}} \text{Vol}(\hat{\sigma})$ . Then Alg. 1 always halts for any  $\varepsilon$  or, in other words,  $\lim_{t \rightarrow \infty} s_t = 0$ .*

We refer to the Appendix for a proof. Although the Lipschitz assumption is necessary for the theoretical proof, we empirically show in the experimental section that the ANNR is well-behaved even for highly discontinuous functions such as characteristic functions of geometrically articulated domains (see Sec. 5.2).

The proof of Proposition 4.2 remains valid when  $s_t$  and the corresponding maximum is computed w.r.t. a partial Delaunay triangulation coming from the approximation discussed in Sec. 3.3. That is, the ANNR is guaranteed to halt even with the approximation procedure.

## 5. Experiments

We select DEFER as our primary baseline as it is the state-of-the-art method for active function approximation which is suitable for arbitrary functions (Bodin et al., 2021). As an ablation, we additionally compare with the non-active version of the NNR which samples datapoints *uniformly* from  $A$ , denoted by nANNR.

In all the experiments, the number of queries is referred to as  $N$ . For numerical validation, we rely on the standard score Mean Average Error  $\text{MAE} = \frac{1}{|P_{\text{test}}|} \sum_{p \in P_{\text{test}}} |\tilde{f}(p) - f(p)|$ , where  $P_{\text{test}}$  is a test set. We generate  $P_{\text{test}}$  as an equally-spaced grid in  $m = 2$  dimensions and by uniformly sampling from  $A$  when  $m > 2$  since exhaustive grid sampling is computationally infeasible.

### 5.1. Hyperparameter $\lambda$

As discussed in Sec. 3.1 and 4.1, the ANNR has a single hyperparameter – the lifting coefficient  $\lambda$  – which governs a natural exploration-exploitation trade-off. This is graphically demonstrated in Fig. 5, where a higher  $\lambda$  encourages querying in areas where  $f$  varies the most. In practice, we suggest the following heuristic choice for  $\lambda$ , which we implement in our experiments. We select  $\lambda$  proportional to the size of the domain and inversely proportional to the scale of the function, effectively bringing domain and codomain to the same scale to balance exploration and exploitation of the function:  $\lambda = \frac{\text{Vol}(A)}{\max f - \min f}$ . Here,  $\max$  and  $\min$  can be estimated from prior knowledge or, more concretely, directly evaluated from the initial dataset  $P_0$ .

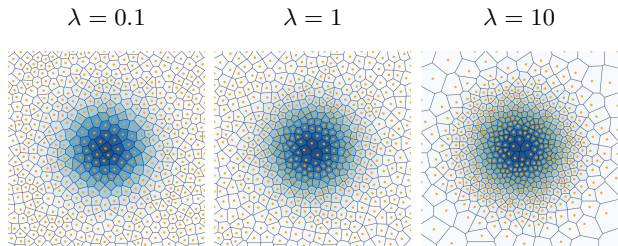


Figure 5: ANNR approximation of a normalized Gaussian with  $\sigma^2 = 0.1$  for various values of  $\lambda$  ( $N = 500$ ).

## 5.2. Geometric Advantages of the ANNR

In this experiment, we consider characteristic functions with various supports and demonstrate that the ANNR is well suited for approximating the support shape. The ANNR leverages the variable geometry of Voronoi cells, which results in a fine-grained approximation with a small number of queries. In contrast, DEFER relies on a rectangular partitioning of the space, which, as we show, can result in sub-optimal approximations. This feature also allows ANNR to approximate functions with arbitrary compact domains, in particular those with a small volume compared to  $A$  (see the Appendix for an example).

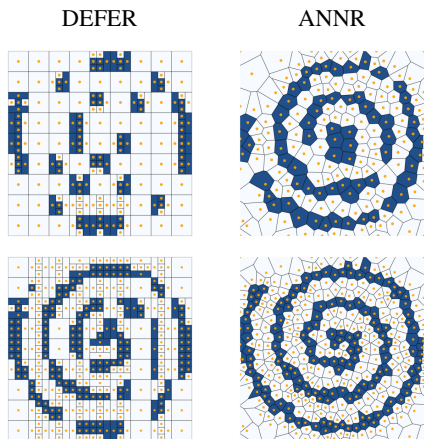


Figure 6: Approximation of a spiral characteristic function (see the Appendix for details).  $N = 200$  (top),  $N = 400$  (bottom).

We first consider the characteristic function of a spiral (Fig. 6). The ANNR displays a visibly better approximation and has a connected support already after  $N = 400$  queries. The support of DEFER is instead highly disconnected as it is unable to capture regions where the spiral is misaligned with the Cartesian axes due to its rectangular bias (see the Appendix for details).

**Rotational Equivariance.** DEFER is sensitive to rotations of the domain due to its inherent bias towards Cartesian axes. In contrast, the ANNR allows to approximate shapes

without any orientation bias since Voronoi cells are arbitrary polytopes. Fig. 7 demonstrates improved stability of the ANNR approximation with respect to rotations of  $\mathbb{R}^2$  compared to DEFER (see the Appendix for details).

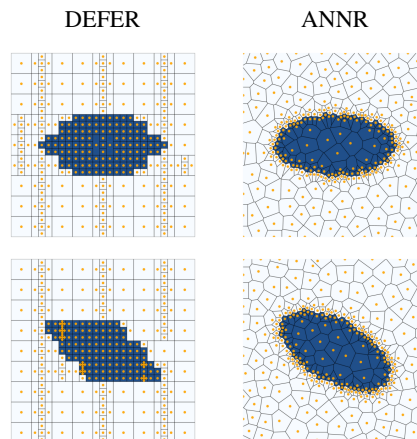
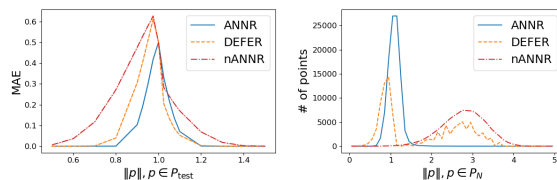


Figure 7: Approximation of an ellipse characteristic function  $\mathbf{1}_{x^2+4y^2 \leq 1}$  with  $0^\circ$  (top) and  $30^\circ$  (bottom) rotations of the domain ( $N = 300$ ).

**Curse of Dimensionality.** The bias of DEFER towards rectangular geometry can potentially affect the quality of approximation in high-dimensions due to the increasing *spikiness* of rectangles (Devi et al., 2016; Blum et al., 2020). To demonstrate this, we consider a characteristic function  $\mathbf{1}_{\|x\| \leq 1}$  of a unit ball in  $\mathbb{R}^6$  (with  $A = [-2, 2]^6$ ). Fig. 8 reports the score (Fig. 8a) and query point distribution (Fig. 8b) with respect to the norm of test and query points respectively.



(a) MAE scores for test sets at different norms. (b) Norm distribution of the queried points.

Figure 8: Analysis of the 6-dimensional unit ball approximations based on the norm of data ( $N = 10^5$  and  $|P_{\text{test}}| = 10^7$  per norm).

Fig. 8a shows that the MAE scores decrease when the test points are sampled away from the discontinuity of  $f$  i.e., the boundary of the ball. ANNR’s score is significantly lower inside the ball ( $\|p\| < 1$ ,  $p \in P_{\text{test}}$ ). At the same time, Fig. 8b shows that all the queries of the ANNR are queried in the proximity of the boundary, while DEFER’s query norm distribution is shifted towards the uniform one (nANNR).

### 5.3. Performance Comparison

In this section we compare the performance of the ANNR with DEFER and the non-active sampler nANNR for articulated functions including parameter estimation of gravitational waves and exploration of the latent space of a deep generative model. The scores are presented in Table 1. For the ANNR and the nANNR, we report averages and standard deviations over 10 runs with different random initializations (in contrast, DEFER has no stochasticity). We set  $N = 1000$  for the two-dimensional experiment (latent manifold exploration) and  $N = 10^5$  when  $m > 2$ . Our method outperforms the baseline in terms of MAE by a significant margin.

Table 1: Performance Comparison (MAE).

	ANNR	DEFER	nANNR
Gravitational waves parameter estimation	0.4710 $\pm 0.0232$	0.5309	0.5317 $\pm 0.2175$
Latent space volume density	47.0509 $\pm 0.4975$	50.5073	54.6290 $\pm 0.8010$

**Gravitational Waves.** A large area of application for active function approximation is astrophysics and, in particular, gravitational wave parameter inference problems (Speagle, 2020; Ashton et al., 2019). Observation of gravitational waves are rare, and the magnitude of the effect is low, implying scarcity of collected data. With limited observations and dozens of parameters describing the gravitational event, the problem of parameter estimation is commonly solved by Bayesian inference methods approximating a simulated log-likelihood function on the parameters (Ashton et al., 2019; Thrane & Talbot, 2019).

We use the same 6-dimensional formulation of parameter inference as described in (Bodin et al., 2021), and refer the reader to the Appendix for a detailed description of parameters. In addition to the original setup, we rescale the function domain to a unit hypercube  $A = [0, 1]^6$  for convenience and multiply  $f$  by a factor of  $e^{8000}$  for better numerical stability. Lastly, in order to deal with practical unboundedness of the density function, we perform an adaptive clipping of extensively sharp volumes. The details of the clipping are available in the Appendix.

Apart from the MAE comparison in Table 1, we present a visualization of the function approximations in Fig. 9 of a single run. The figure displays histograms over marginal distributions over a set of two-dimensional slices of the domain. Each marginalization was performed via Monte-Carlo sampling with  $10^5$  samples per bin. The figure shows the visual closeness of ANNR marginalizations to the original distribution compared to the baseline. Marginals over all 15 pairs of parameters are available in the Appendix.

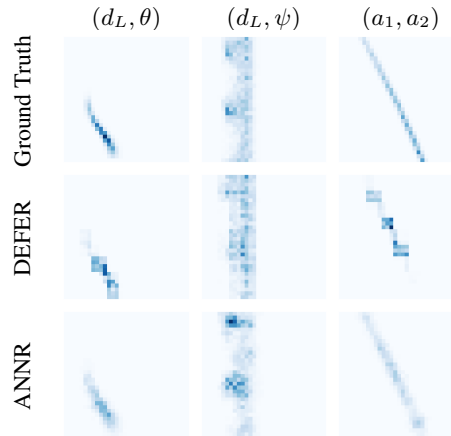


Figure 9: Marginal distributions for gravitational waves over selected two-dimensional slices. Each slice is identified by a pair of parameters named as in (Bodin et al., 2021).

**Latent Manifold Exploration.** Our last experiment addresses a manifold exploration task on the latent space of a deep generative model. The generative process enables to query an arbitrary point in the latent space and thus suits the framework of the ANNR and DEFER. Moreover, decoding the queried points allows to interpret and semantically evaluate the exploration procedure. To this end, we deploy a generative model  $\varphi : \mathcal{Z} \rightarrow \mathcal{X}$  that maps a prior distribution on the latent space  $\mathcal{Z} = \mathbb{R}^m$  to the data distribution on  $\mathcal{X} = \mathbb{R}^n$  and regard it as (the probabilistic analogue of) a parametrization of the data manifold. The ‘mass’ distribution of data on  $\mathcal{Z}$  is represented by the induced *volume density*  $f$  i.e., the density of the volume form of the Riemannian metric induced by  $\varphi$  on  $\mathcal{Z}$  via pull-back (Arvanitidis et al., 2018; 2021). Such volume density is concretely expressed as  $f(z) = \sqrt{\det(J_\varphi(z)^T J_\varphi(z))}$  where  $J_\varphi(z)$  is the Jacobian matrix of  $\varphi$  at  $z \in \mathcal{Z}$ . Intuitively,  $f$  can be seen as a ‘fuzzy’ characteristic function of the latent manifold and an active function approximator for  $f$  can be interpreted as progressively exploring such manifold.

We train  $\varphi$  as (the decoder of) a Variational Autoencoder (VAE, Kingma & Welling (2014)) on the MNIST dataset (Deng, 2012) of gray-scale images of hand-written digits ( $n = 784$ ). We deploy a two-dimensional latent space ( $m = 2$ ) with a standard Gaussian prior. Despite its low dimensionality, such a latent space is sufficient for generation of quality MNIST images (Kingma & Welling, 2014) and enables direct visualization. For better parametrization quality, the latent prior is additionally encouraged by a hyperparameter  $\beta = 2$  multiplying the corresponding ELBO loss term (Higgins et al., 2017).

Fig. 10 displays the progressive approximation of  $f$  as well as images decoded from queried points. As can be seen from the latter, around  $t = 500$  the class of  $p_t$  (digit) is



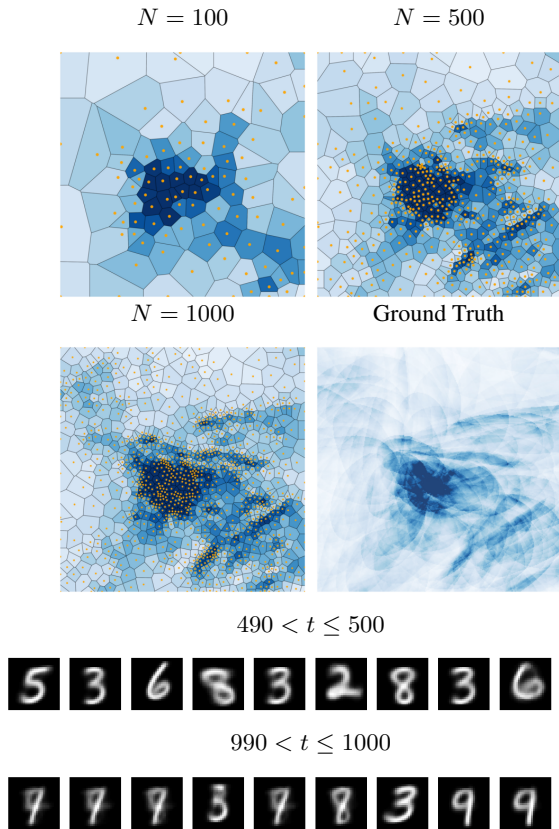


Figure 10: **Top:** approximation of the latent space volume density of a deep generative model. **Bottom:** images decoded from the queried latent points  $p_t$  for two intervals of  $t$ .

different at each step  $t$  and the ANNR is thus covering a wide semantic range of data. In contrast, around  $t = 1000$  the decoded images are similar, indicating a phase close to convergence in which the approximation is refined locally.

**Runtime Comparison.** The overall flexibility of the geometric framework utilized by the ANNR comes with a certain computational cost, which highlights a tradeoff between effectiveness and efficiency of the methods. The computational cost of the ANNR is largely mitigated by the Delaunay approximation the method relies on. In addition, the runtime of the ANNR can be heavily controlled by adjusting the desired Markov Chain sampling precision and increasing the number of available parallel threads.

We provide a runtime comparison between the methods on experiments discussed earlier in this section in Table 2. Each number represents the average runtime for a single experimental run. All experiments are performed on CPU Ryzen 9 5950X 16-Core. We note that in case of the ANNR single queries still take milliseconds to compute, which is reasonable for real-life applications with high function

evaluation cost.

Table 2: Runtime Comparison.

	ANNR	DEFER
Gravitational waves parameter estimation	1347 s	186 s
Latent space volume density	220 ms	90 ms

## 6. Conclusion and Future Work

In this work we introduced the ANNR – an adaptation of nearest neighbor regression to an active setting. We provided a computationally efficient implementation of the ANNR as well as theoretical halting guarantees. Our empirical investigations have shown that the ANNR outperforms the state-of-the-art active function approximator called DEFER.

An interesting line for future investigation lies in designing active querying strategies for other higher-dimensional function approximators from the existing literature. Examples include the piecewise linear Delaunay interpolator (Omohundro, 1989) or the tensor product of cubic splines (Lalescu, 2009).

## 7. ACKNOWLEDGEMENTS

This work was supported by the Swedish Research Council, the Knut and Alice Wallenberg Foundation and the European Research Council (ERC-BIRD-884807).

## References

- Acerbi, L. Variational bayesian monte carlo. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18*, pp. 8223–8233, Red Hook, NY, USA, 2018. Curran Associates Inc.
- Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.
- Arvanitidis, G., Hansen, L. K., and Hauberg, S. Latent space oddity: on the curvature of deep generative models. In *6th International Conference on Learning Representations, ICLR 2018, Conference Track Proceedings*, 2018.
- Arvanitidis, G., Hauberg, S., and Schölkopf, B. Geometrically enriched latent spaces. In Banerjee, A. and Fukumizu, K. (eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 631–639. PMLR, 13–15 Apr 2021.
- Ashton, G., Hübner, M., Lasky, P. D., Talbot, C., Ackley, K., Biscoveanu, S., Chu, Q., Divakarla, A., Easter, P. J., Goncharov, B., Vivanco, F. H., Harms, J., Lower, M. E., Meadors, G. D., Melchor, D., Payne, E., Pitkin, M. D., Powell, J., Sarin, N., Smith, R. J. E., and Thrane, E. Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *The Astrophysical Journal Supplement Series*, 241(2):27, 2019.
- Baum, E. B. and Lang, K. Query learning can work poorly when a human oracle is used. In *International joint conference on neural networks*, volume 8, pp. 8, 1992.
- Bentley, J. L. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.
- Bern, M. and Eppstein, D. Mesh generation and optimal triangulation. *Computing in Euclidean geometry*, 1:23–90, 1992.
- Blum, A., Hopcroft, J., and Kannan, R. *Foundations of data science*. Cambridge University Press, 2020.
- Bodin, E., Dai, Z., Campbell, N., and Ek, C. H. Black-box density function estimation using recursive partitioning. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1015–1025. PMLR, 18–24 Jul 2021.
- Bondu, A., Lemaire, V., and Boullé, M. Exploration vs. exploitation in active learning : A bayesian approach. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, 2010.
- Brezis, H. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.
- Chen, L. Mesh smoothing schemes based on optimal delaunay triangulations. In *IMR*, pp. 109–120. Citeseer, 2004.
- Chew, L. P. Guaranteed-quality mesh generation for curved surfaces. In *Proceedings of the ninth annual symposium on Computational geometry*, pp. 274–280, 1993.
- Cohn, D., Atlas, L., and Ladner, R. Improving generalization with active learning. *Machine learning*, 15(2): 201–221, 1994.
- Delaunay, B. et al. Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk*, 7(793-800):1–2, 1934.
- Deng, L. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.
- Devi, G., Chauhan, C., and Chakraborti, S. Conceptualizing curse of dimensionality with parallel coordinates. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- Devillers, O., Pion, S., and Teillaud, M. Walking in a triangulation. In *Proceedings of the seventeenth annual symposium on Computational geometry*, pp. 106–114, 2001.
- Edelsbrunner, H. and Harer, J. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- Fortune, S. Voronoi diagrams and delaunay triangulations. *Computing in Euclidean geometry*, pp. 225–265, 1995.
- Gubaev, K., Podryabinkin, E. V., and Shapeev, A. V. Machine learning of molecular properties: Locality and active learning. *The Journal of chemical physics*, 148(24): 241727, 2018.
- Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A. beta-vae: Learning basic visual concepts with a constrained variational framework. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.
- Higson, E., Handley, W., Hobson, M. P., and Lasenby, A. N. Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation. *Statistics and Computing*, pp. 1–23, 2019.

- Järvenpää, M., Gutmann, M., Vehtari, A., and Marttinen, P. Efficient acquisition rules for model-based approximate bayesian computation. *Bayesian Analysis*, 14, 06 2019.
- Järvenpää, M., Vehtari, A., and Marttinen, P. Batch simulations and uncertainty quantification in gaussian process surrogate approximate bayesian computation. In Peters, J. and Sontag, D. (eds.), *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pp. 779–788. PMLR, 03–06 Aug 2020.
- Kandasamy, K., Schneider, J., and Póczos, B. Bayesian active learning for posterior estimation. In *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15*, pp. 3605–3611. AAAI Press, 2015. ISBN 9781577357384.
- Kimberling, C. *Triangle centers and central triangles*. Utilitas Mathematica Publishing Incorporated, 1998.
- Kingma, D. P. and Welling, M. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.
- Klee, V. On the complexity of d-dimensional voronoi diagrams. Technical report, Washington Univ. Seattle Dept. of Mathematics, 1979.
- Kontorovich, A., Sabato, S., and Uerner, R. Active nearest-neighbor learning in metric spaces. In *Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16*, pp. 856–864, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.
- Lalescu, C. C. Two hierarchies of spline interpolations. practical algorithms for multivariate higher order splines. *CoRR*, abs/0905.3564, 2009.
- Lewis, D. D. and Gale, W. A. A sequential algorithm for training text classifiers. In *SIGIR'94*, pp. 3–12. Springer, 1994.
- Liu, H., Ong, Y.-S., Shen, X., and Cai, J. When gaussian process meets big data: A review of scalable gps. *IEEE Transactions on Neural Networks and Learning Systems*, 31(11):4405–4423, 2020.
- MacKay, D. J. C. Information-based objective functions for active data selection. *Neural Computation*, 4(4):590–604, 1992.
- Omohundro, S. M. *The Delaunay triangulation and function learning*. International Computer Science Institute, 1989.
- Polianskii, V. and Pokorný, F. T. Voronoi graph traversal in high dimensions with applications to topological data analysis and piecewise linear interpolation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2154–2164, 2020.
- Ramakrishnan, R., Dral, P. O., Rupp, M., and von Lilienfeld, O. A. Big data meets quantum chemistry approximations: the  $\delta$ -machine learning approach. *Journal of chemical theory and computation*, 11(5):2087–2096, 2015.
- Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B., Chen, X., and Wang, X. A survey of deep active learning. *ACM Computing Surveys (CSUR)*, 54(9):1–40, 2021.
- Ruppert, J. A delaunay refinement algorithm for quality 2-dimensional mesh generation. *Journal of algorithms*, 18(3):548–585, 1995.
- Rushdi, A., Swiler, L. P., Phipps, E. T., D’Elia, M., and Ebeida, M. S. Vps: Voronoi piecewise surrogate models for high-dimensional data fitting. *International Journal for Uncertainty Quantification*, 7(1), 2017.
- Settles, B. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
- Shewchuk, J. R. Delaunay refinement algorithms for triangular mesh generation. *Computational geometry*, 22 (1-3):21–74, 2002.
- Sommerville, D. M. Y. *An Introduction to the Geometry of n Dimensions*, volume 512. Dover New York, 1958.
- Speagle, J. S. dynasty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society*, 493 (3):3132–3158, 02 2020. ISSN 0035-8711.
- Sung, K. and Niyogi, P. Active learning for function approximation. In Tesauro, G., Touretzky, D., and Leen, T. (eds.), *Advances in Neural Information Processing Systems*, volume 7, pp. 593–600. MIT Press, 1995.
- Thrane, E. and Talbot, C. An introduction to bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. *Publications of the Astronomical Society of Australia*, 36:e010, 2019.
- Zhu, J.-J. and Bento, J. Generative adversarial active learning. *arXiv preprint arXiv:1702.07956*, 2017.

## A. Proofs of Theoretical Results

In this section we provide proofs for the theoretical results presented in the main body of the paper. We start by proving the result from Sec. 4.1.

**Proposition A.1.** *There exists a constant  $C > 0$  such that for every smooth function  $f : \Omega \rightarrow \mathbb{R}$  the following inequality holds:*

$$\log \text{Vol}(\Gamma_f) \geq C \|f - f_\Omega\|_2^2 + \log \text{Vol}(\Omega) + o(\|g_f\|^2). \quad (8)$$

*Proof.* The volume of  $\Gamma_f$  can be expressed as

$$\text{Vol}(\Gamma_f) = \int_{\Omega} \sqrt{\det(1 + g_f)} \quad (9)$$

and thus by Jensen inequality we get:

$$\log \frac{\text{Vol}(\Gamma_f)}{\text{Vol}(\Omega)} \geq \frac{1}{2\text{Vol}(\Omega)} \int_{\Omega} \log \det(1 + g_f). \quad (10)$$

Since by general properties of matrices  $\log \det(g_f) = \text{tr}(\log g_f)$  and  $\log(1 + g_f) = g_f + o(\|g_f\|^2)$  where  $\|\cdot\|$  denotes the standard matrix norm, the right hand side of Eq. 10 reduces to

$$C_1 \int_{\Omega} \text{tr}(g_f) + o(\|g_f\|^2) = C_1 \|\nabla_f\|_2^2 + o(\|g_f\|^2). \quad (11)$$

Finally, by Poinaré-Wirtinger inequality (Brezis, 2010)  $\|\nabla_f\|_2^2 \geq C_2 \|f - f_\Omega\|_2^2$  for some constant  $C_2 > 0$ , which concludes the proof.  $\square$

We now prove the halting guarantee from Sec. 4.2. First, we provide the following result concerning high-dimensional Euclidean geometry which will be necessary for the main proof. Recall that  $A \subseteq \mathbb{R}^m$  is a convex compact set.

**Lemma A.2.** *For each  $\delta > 0$  there exists an  $\eta > 0$  such that for each  $n$ -dimensional simplex  $\sigma \subseteq A$  if  $\text{Vol}(\sigma) > \delta$  then  $d(v, x) > \eta$  for each vertex  $v$  of  $\sigma$  and each point  $x$  of the segment connecting the baricenter and the circumcenter of  $\sigma$ .*

*Proof.* By comparing the volume of a simplex with the volume of its circumsphere, there exists  $\eta_1 > 0$  such that for any simplex  $\sigma$  if  $\text{Vol}(\sigma) > \delta$  then  $R_\sigma > \eta_1$ , where  $R_\sigma$  denotes the radius of the circumsphere of  $\sigma$ . There also exists a  $\eta_2 > 0$  such that for any simplex  $\sigma$  if  $\text{Vol}(\sigma) > \delta$  then all the heights (i.e., the segments passing through the vertices and orthogonal to the opposite faces) of  $\sigma$  are greater than  $\eta_2$ . To see this, note that otherwise there would exist triangles of arbitrarily small heights and thus, by the volume constraint, with faces of arbitrarily large volume, contradicting the compactness of  $A$ . Now, given a simplex  $\sigma$  and a vertex  $v$  of  $\sigma$ , the segment connecting  $v$  to its opposite face and containing the barycenter  $b$  of  $\sigma$  is shorter than the height passing through  $v$  because of orthogonality of the latter. The barycenter separates that segment into a portion of  $\frac{m-1}{m}$  of it containing  $v$ . Thus, we have  $d(v, b) > \frac{m-1}{m} \eta_2$ .

Consider  $\eta_3 = \min\{\eta_1, \frac{m-1}{m} \eta_2\}$ . For any simplex  $\sigma$  with circumcenter  $c$  and barycenter  $b$  we then have  $d(v, b) > \eta_3$  and  $d(v, c) > \eta_3$  for every vertex  $v$  of  $\sigma$ . Since  $(b - v) \cdot (c - v) > 0$ , we have that  $d(v, x) > \frac{\sqrt{2}}{2} \eta_3 = \eta$  for each  $x$  in the segment connecting  $b$  and  $c$ , as desired.  $\square$

We now prove the main result from Sec. 4.2.

**Proposition A.3.** *Assume that the ground-truth function  $f$  is Lipschitz and let  $s_t = \max_{\sigma \in \text{Del}_{P_t}} \text{Vol}(\hat{\sigma})$ . Then Alg. 1 always halts for any  $\varepsilon$  or, in other words,  $\lim_{t \rightarrow \infty} s_t = 0$ .*

*Proof.* Suppose by contradiction that there exists  $\varepsilon > 0$  such that  $s_t > \varepsilon$  for all sufficiently large  $t$ . Since  $f$  (and thus  $\lambda f$ ) is Lipschitz, there exists  $\delta > 0$  such that for any simplex  $\sigma$  if  $\text{Vol}(\hat{\sigma}) > \varepsilon$  then  $\text{Vol}(\sigma) > \delta$ .

Let us prove that there can not be infinite queries in the interior of the bounding region  $A$ . By comparing the volume of a simplex with the volume of its circumsphere, there exists  $\gamma > 0$  such that for any simplex  $\sigma$  if  $\text{Vol}(\sigma) > \delta$  then  $R_\sigma > \gamma$ , where  $R_\sigma$  denotes the radius of the circumsphere of  $\sigma$ . When a point  $p_{t+1} = \text{Circ}(\bar{\sigma})$  is queried, since by hypothesis

$s_t = \text{Vol}(\hat{\sigma}) > \epsilon$  we have  $R_{\hat{\sigma}} > \gamma$ . But then  $d(p_{t+1}, q) > \gamma$  for all  $q \in P_t$  since no points in  $P_t$  are contained inside the circumsphere of the Delaunay simplex  $\hat{\sigma}$ . If infinite points in the interior of  $A$  are queried, one gets an infinite sequence in  $A$  whose pairwise distances are all greater than  $\gamma$ . This is impossible since  $A$  is compact.

As a consequence, the queried points  $p_{t+1}$  belong to the boundary of  $A$  for  $t \gg 0$  and are obtained via the method described in Sec. 3.2. By Lemma A.2 there exists  $\eta > 0$  such that  $d(v, p_{t+1}) > \eta$  for  $t \gg 0$  where  $v$  is any vertex of the simplex  $\bar{\sigma}$  whose lifting has the largest volume at step  $t$ . Since  $\bar{\sigma}$  is a Delaunay simplex it holds that  $d(q, p_{t+1}) > \eta$  for every  $q \in P_t$  and  $t \gg 0$ , which again contradicts the compactness of  $A$ .  $\square$

## B. Extended Experiments

### B.1. Spiral

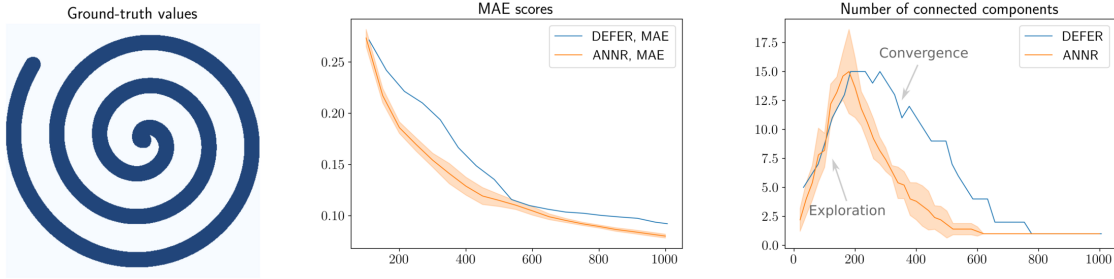


Figure 11: Ground-truth plot of a spiral characteristic function and performance comparison between the ANNR and DEFER as the number of queries increases. Scores are averaged over 5 runs.

### B.2. Arbitrary Domain

In many cases the actual domain of a function is not the entire  $\mathbb{R}^m$ , but rather some compact subspace. In many optimization algorithms, the function domain is artificially extended to a bounding volume (usually a bounding box) containing it, assuming some zero value outside of the original function domain. In higher dimensions, this significantly increases the volume to explore (Devi et al., 2016). The ANNR takes advantage of the flexibility of Delaunay partitioning to restrict new queries to the (boundary of the) function domain. Fig. 12 illustrates an approximation of  $f(x) = \|x\|$  restricted to the intersection of two circles (the volume of the domain is only  $\sim 1/30$  of the bounding box volume), which otherwise can be very inefficiently approximated by its bounding box.

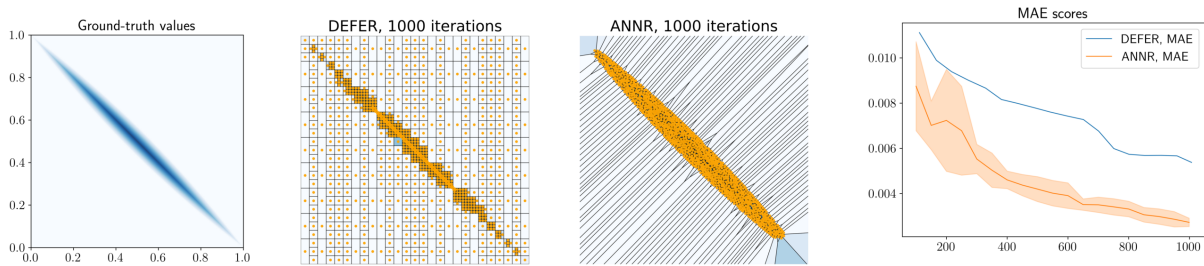


Figure 12: Approximation of a distance function defined on an intersection of two circles with radius 5 and centered at  $(-3, -3)$  and  $(4, 4)$ .

### B.3. Rotational Equivariance

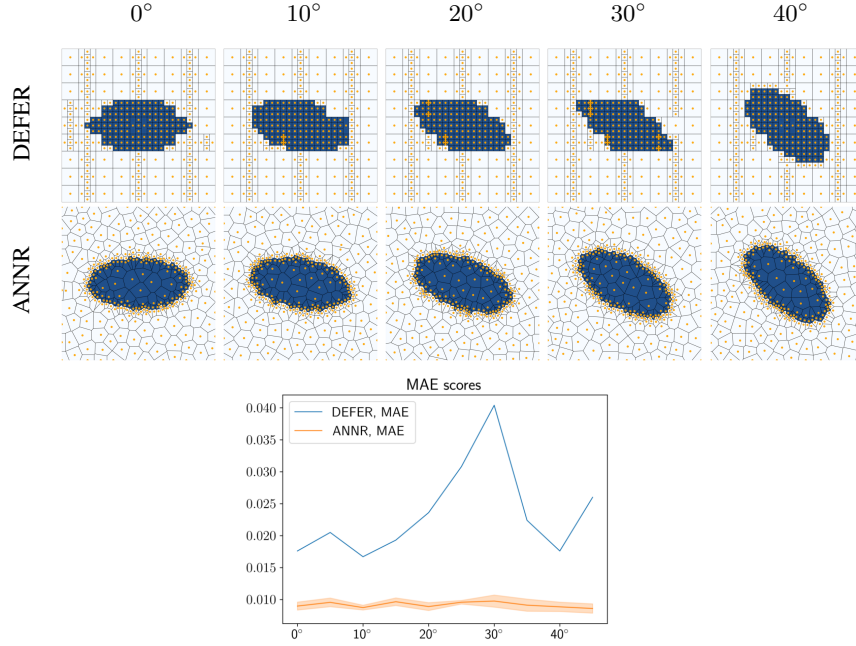


Figure 13: **Top**: approximation of an ellipse characteristic function  $\mathbf{1}_{x^2+4y^2 \leq 1}$  by DEFER and the ANNR with various rotations of the domain ( $N = 300$ ). **Bottom**: performance comparison as the angle varies.

### B.4. Gravitational Waves

**Data preprocessing.** We follow the initial setup of the problem described in (Bodin et al., 2021). Namely, the function is constructed from a gravitational wave example provided [tutorial](#). The six parameters of a gravitational wave generated by a binary black hole that are inferred by the model are: the luminosity distance  $d_L$ , the inclination angle  $\iota$ , the polarization angle  $\psi$ , the phase  $\phi_c$  relative to a reference time and two spin magnitudes  $a_1$  and  $a_2$ . For more details about the nature of the data we refer to (Ashton et al., 2019).

Evaluation of the logarithm of the approximated likelihood on several uniformly selected points in the domain yields values distributed around  $-8000$  (the simulator provides log-likelihood). Since our aim is to approximate the original likelihood function, we add 8000 to all log-density evaluations and then exponentiate the values, effectively multiplying the likelihood function by  $e^{8000}$ . This operation brings the majority of output values of the underlying function close to single digits, stabilizing the computations. Such scaling does not affect the baseline as DEFER is invariant to such transformations, and multiplicatively affects our choice of the lifting parameter.

**Volume clipping.** While the approximated density function may appear to take relatively low values over the large part of the domain, some of the concentrated areas may produce values many orders of magnitude higher than that. Without any Lipschitz guarantees for the underlying function, such areas could create attractors for the ANNR, forcing the method’s exploration to ‘sink’ in such singularities and over-exploit the area. In order to mitigate that, we propose to truncate the scores of simplices in accordance to a pre-selected sensible Lipschitz constant.

Consider a simplex  $\sigma \in \text{Del}_P$  and its lifting  $\hat{\sigma}$  and note that  $\text{Vol}(\hat{\sigma}) = \frac{1}{\cos \alpha} \text{Vol}(\sigma)$ , where  $\alpha$  is the *dihedral angle* between  $\sigma$  and  $\hat{\sigma}$  both naturally embedded in  $\mathbb{R}^{m+1}$ . Limiting the Lipschitz constant is equivalent to limiting the maximal dihedral angle to some  $\alpha_0$ . As the result of the clipping,  $\text{Vol}(\hat{\sigma})$  in Alg. 1 gets transformed into  $\min\{\text{Vol}(\hat{\sigma}), \frac{1}{\cos \alpha_0} \text{Vol}(\sigma)\}$ . In our experiments with gravitational waves, we use  $\alpha_0 = 89^\circ$ .

**Marginals.** Fig. 14 presents the marginalizations of the approximated function over all possible two-parameter slices.

### Active Nearest Neighbor Regression Through Delaunay Refinement

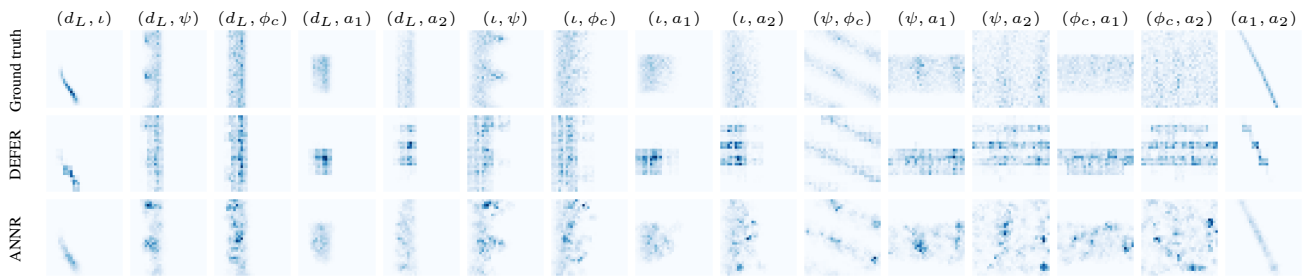


Figure 14: Marginal distribution for gravitational waves over all two-dimensional slices of parameters. Each slice is identified by a pair of parameters named as in (Bodin et al., 2021).