Weighted ε-Nets

Daniel Bertschinger[1] and Patrick Schnider[2]

1 **Department of Computer Science, ETH Zürich, Switzerland**
`daniel.bertschinger@inf.ethz.ch`
2 **Department of Computer Science, ETH Zürich, Switzerland**
`patrick.schnider@inf.ethz.ch`

Abstract

Motivated by recent work of Bukh and Nivasch [4] on one-sided ε-approximants, we introduce the notion of *weighted ε-nets*. It is a geometric notion of approximation for point sets in $\mathbb{R}^d$ similar to ε-nets and ε-approximations, where it is stronger than the former and weaker than the latter. The main idea is that small sets can contain many points, whereas large sets must contain many points of the weighted ε-net.

In this paper, we analyze weak weighted ε-nets with respect to convex sets and axis-parallel boxes and give upper and lower bounds on ε for weighted ε-nets of size two and three. Some of these bounds apply to classical ε-nets as well.

1 Introduction

Representing large, complicated objects by smaller, simpler ones is a common theme in mathematics. For one-dimensional data sets this is realized by the notions of medians, means and quantiles. One fundamental difference between medians and quantiles on the one side and the mean on the other side is the robustness of the former against outliners of the data.

Centerpoint. Medians and quantiles are one-dimensional concepts, whereas modern data sets are often multidimensional. Hence, many generalizations of medians and quantiles to higher dimensions have been introduced and studied. One example is the notion of a centerpoint, that is, a point c such that for every closed halfspace h containing c we know that h contains at least a $\frac{1}{d+1}$-fraction of the whole data, where d denotes the dimension. The Centerpoint Theorem ensures that for any point set in $\mathbb{R}^d$ there always exists such a centerpoint [11].

Instead of representing a data set by a single point, one could take a different point set as a representative. This is exactly the idea of an *ε-net*.

▶ **Definition 1.1.** Given any range space $(X, \mathcal{R})$, an *ε-net* on a point set $P \subseteq X$ is a subset $N \subseteq P$ such that every $R \in \mathcal{R}$ with $|R \cap P| \geq \varepsilon|P|$ has nonempty intersection with N. If the condition that an ε-net needs to be a subset of P is dropped, then N is called a *weak ε-net*.

In this language, a centerpoint is a weak $\frac{d}{d+1}$-net for the range space of halfspaces. The concept of ε-nets has been studied in a huge variety; first, there are statements on the existence and the size of ε-nets, if ε is given beforehand. On the other hand, one can fix the size of the ε-net a priori and try to bound the range of ε in which there always exists an ε-net. For the former, it is known that every range space of VC-dimension δ has an ε-net of size at most $\mathcal{O}(\frac{\delta}{\varepsilon}\log\frac{1}{\varepsilon})$ [7].

ε-Approximations. For some applications though, ε-nets may not retain enough information. For every range we only know that it has a nonempty intersection with the net; however, we do not know anything about the size of this intersection. Hence, the following definition of ε-approximations comes naturally.

36th European Workshop on Computational Geometry, Würzburg, Germany, March 16–18, 2020.
This is an extended abstract of a presentation given at EuroCG'20. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear eventually in more final form at a conference with formal proceedings and/or in a journal.

▶ **Definition 1.2.** Given any range space $(X, \mathcal{R})$ and any parameter $0 \leq \varepsilon \leq 1$, an *$\varepsilon$-approximation* on a point set $P \subset X$ is a subset $A \subset P$ such that for every $R \in \mathcal{R}$ we have $\left|\frac{|R \cap P|}{|P|} - \frac{|R \cap A|}{|A|}\right| \leq \varepsilon$.

Initiated by the work of Vapnik and Chervonenkis [13], one general idea is to construct ε-approximations by uniformly sampling a random subset $A \subseteq X$ of large enough size. This results in statements about the existence of ε-approximations depending on the VC-dimension of the range space. In particular every range space of VC-dimension δ allows an ε-approximation of size $\mathcal{O}(\frac{\delta}{\varepsilon^2} \log \frac{1}{\varepsilon})$ [5, 6, 8].

Convex Sets. It is well-known that the range space of convex sets has unbounded VC-dimension; therefore, none of the results mentioned above can be applied. While constant size weak ε-nets still exist for the range space of convex sets [1, 12], the same cannot be said for weak ε-approximations (Proposition 1 in [4]). Motivated by this, Bukh and Nivasch [4] introduced the notion of *one-sided weak ε-approximants*. The main idea is that small sets can contain many points, whereas large sets must contain many points of the approximation. Bukh and Nivasch show that constant size one-sided weak ε-approximants exist for the range space of convex sets. In this work, we define a similar concept, called *weighted ε-nets*. In contrast to one-sided weak ε-approximants, our focus is to understand what bounds can be achieved for a fixed small value of k, which is given a priori. In this sense our approach is similar to the one taken by Aronov et al. [2] (for standard ε-nets).

▶ **Definition 1.3.** Given any point set $P \subset \mathbb{R}^d$ of size n, a *weighted ε-net of size k* (with respect to some range space) is defined as a set of points $p_1, \ldots, p_k$ and some values $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k)$ such that every set in the range space containing more than $\varepsilon_i n$ points of P contains at least i of the points $p_1, \ldots, p_k$.

Following historic conventions, we denote a weighted ε-net as *strong* if $p_1, \ldots, p_k \in P$ and as *weak* otherwise. In this work, we focus on weak weighted ε-nets of small size for the range space of convex sets and axis-parallel boxes.

2 Weighted ε-nets for the range space of convex sets

Weighted ε-nets for the range space of halfspaces were already studied by Pilz and Schnider [10]. In this section we generalize one of their results to the range space of convex sets.

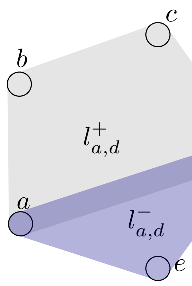
▶ **Theorem 2.1.** *Let P be a set of n points in general position in $\mathbb{R}^d$. Let $0 < \varepsilon_1 \leq \varepsilon_2 < 1$ be arbitrary constants with (i) $d\varepsilon_1 + \varepsilon_2 \geq d$ and (ii) $\varepsilon_1 \geq \frac{2d-1}{2d+1}$. Then there are two points p_1 and p_2 in $\mathbb{R}^d$ such that*
1. *every convex set containing more than $\varepsilon_1 n$ points of P contains at least one of the points p_1 or p_2, and*
2. *every convex set containing more than $\varepsilon_2 n$ points of P contains both p_1 and p_2.*

In the following we briefly sketch the proof. For a full proof, we refer the interested reader to the full version of this paper [3].

Sketch of Proof. The main idea is to create two classes $\mathcal{A}$ and $\mathcal{B}$, containing convex subsets of $\mathbb{R}^d$. We put every convex subset of $\mathbb{R}^d$ containing more than $\varepsilon_2 n$ points of P (denoted as *big sets*) into both classes $\mathcal{A}$ and $\mathcal{B}$. Further, we put every convex subset of $\mathbb{R}^d$ containing more than $\varepsilon_1 n$ points of P (called the *small sets*) into one of the classes $\mathcal{A}$ or $\mathcal{B}$. To this end we halve P with a $(d-1)$-dimensional hyperplane H. Every small set containing more points of P below H than above H is put into $\mathcal{B}$. Every small set which is not in $\mathcal{B}$, is put into $\mathcal{A}$. It can now be shown that $\mathcal{A}$ as well as $\mathcal{B}$ satisfies the Helly property. We then define p_1 and p_2 as the two Helly points. ◀

3 Lower Bounds on ε

Having seen an existential result for weighted ε-nets with respect to convex sets, we are interested in the best possible value for ε. In this chapter we present some lower bounds on ε. First, an example given in [10] can be adapted to show that inequality (i) of Theorem 2.1 is needed in the following sense: In the plane we cannot simultaneously have $\varepsilon_1 > \frac{3}{5}$ and $\varepsilon_2 > \frac{4}{5}$. To see this, consider the point set in Figure 1. Note that one of the two points needs to lie in $l^+_{a,d} \cap l^-_{a,d}$. The same is true for all intersections depicted in the right part of Figure 1. However, these five intersections cannot be stabbed using only two points.



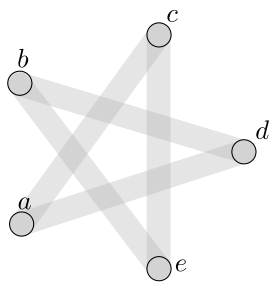


Figure 1 A point set of five regions in convex position, each containing exactly k points. Two particular regions containing four (three, respectively) of the regions. The intersections of interest are drawn on the right side.

3.1 Lower bounds on ε_1

On the other hand, one can give lower bounds on ε_1, independently of the value of ε_2. This setting is exactly the same as giving lower bounds on ε for any ε-net. Hence, any bound given in this chapter is also a lower bound on ε for ε-nets as well. Mustafa and Ray [9] have studied this in dimension 2, showing that there exist point sets P in $\mathbb{R}^2$ such that for every two points p_1 and p_2, not necessarily in P, we can find a convex set containing at least $\frac{4n}{7}$ points of P but neither p_1 nor p_2.

For higher dimension, to our knowledge the bounds given here are among the first and currently the best lower bounds for the range space of convex sets.

▶ **Lemma 3.1.** *There are point sets $P \subset \mathbb{R}^3$, such that for any two points p_1 and p_2 in $\mathbb{R}^3$ we can always find a compact convex set containing at least $\frac{5n}{8}$ points of P, but neither p_1 nor p_2.*

Sketch of Proof. Consider a point set in three dimensions consisting of eight points. There is a hexagon in the xy-plane, one point above the hexagon (denoted as u_1), and one point below the hexagon (denoted as u_2), see Figure 2.

It can be observed that every set of four points of the hexagon and every set of three points together with the center of the hexagon (indicated by the cross) should contain one of p_1 and p_2 for the Lemma to be wrong. However, it is not possible to place p_1 and p_2 accordingly. For more detail we again refer to [3]. ◀

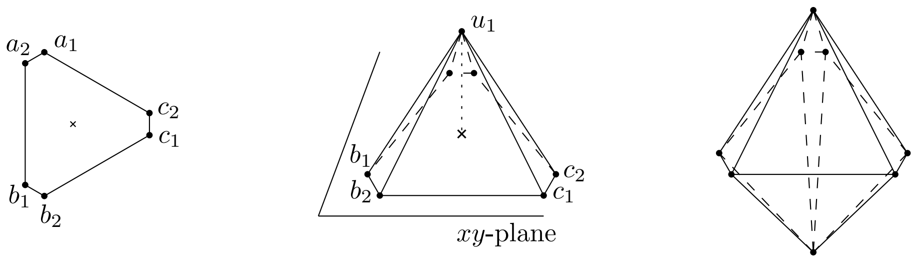


Figure 2 A point set in three dimensions, with six points in the xy-plane arranged in a hexagon, one point above the xy-plane and one point below the xy-plane.

For general dimensions a lower bound on ε_1 is given in the following Lemma. The corresponding examples for the proof consist of a $(d-1)$-dimensional simplex $\mathcal{S}$ in $\mathbb{R}^d$ with exactly one point in every 0-face of the simplex. There are two additional points, one *above* and one *below* the simplex, where above and below refer to the dimension not used for $\mathcal{S}$. A detailed discussion of the arguments can be found in [3].

▶ **Lemma 3.2.** *There are point sets P in $\mathbb{R}^d$ such that for any two points $p_1, p_2 \in \mathbb{R}^d$ there is a compact convex set containing $\frac{d}{d+2}$ of the points of P, but neither p_1 nor p_2.*

4 The range space of axis-parallel boxes

In this section, we study weighted ε-nets of size 2 and 3 for the range space of axis-parallel boxes. Axis-parallel boxes have the property that they allow a much stronger Helly-type result.

▶ **Observation 4.1.** *Let $\mathcal{F}$ be a family of compact, axis-parallel boxes in $\mathbb{R}^d$ such that any two of them have a common intersection. Then the whole collection has a nonempty intersection.*

As a direct consequence of this observation we note that for any point set P in $\mathbb{R}^d$, there always exists a (weighted) $\frac{1}{2}$-net of size 1 for the range space of axis-parallel boxes. For weighted ε-nets of larger size we find the following.

▶ **Theorem 4.2.** *Let P be a set of n points in general position in $\mathbb{R}^d$. Let $0 < \varepsilon_1 \leq \varepsilon_2 < 1$ be arbitrary constants with (i) $\varepsilon_1 \geq \frac{3^{d-1}}{2 \cdot 3^{d-1}+1}$ and (ii) $\varepsilon_1 + \varepsilon_2 \geq 1$. Then there exist two points p_1 and p_2 such that*

1. *every axis-parallel box containing more than $\varepsilon_1 n$ points of P contains at least one of the points p_1 and p_2, and*
2. *every axis-parallel box containing more than $\varepsilon_2 n$ points of P contains both, p_1 and p_2.*

Sketch of Proof. For the sake of simplicity, we only present a proof in $\mathbb{R}^2$ with fixed values $\varepsilon_1 = \frac{3}{7}$ and $\varepsilon_2 = \frac{4}{7}$. For other values the proof works analogously. First, divide the point set with a horizontal line l_1, such that there are $\frac{3n}{7}$ points below l_1 and $\frac{4n}{7}$ points above l_1. Then add two lines l', l'' perpendicular to l_1 splitting the point set below l_1 into three parts containing the same number of points, see Figure 3 (left).

Now one of the two outside areas above l_1, without loss of generality B_1, contains at most $\frac{2n}{7}$ points of P. We then move l' slightly towards l'', until we have the same number of points in B_1' as in A'. We now define $p_1 := l_1 \cap l_2$.

As the area left of l_2 and the area below l_1 contain $\frac{3n}{7}$ of the points of P every big box contains p_1 for sure. On the other hand every small box not containing p_1 lies completely

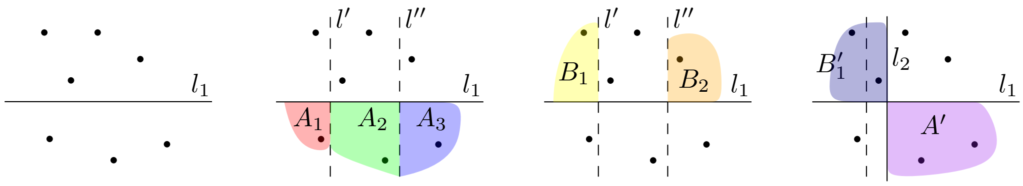


Figure 3 An example of the construction of p_1. First the point set P is split by a line l_1. Then the lines l' and l'' split the point set below l_1 into three disjoint parts containing the same number of points, namely A_1, A_2 and A_3. One of B_1 and B_2 has to contain "few" points of P, without loss of generality B_1, and by slightly changing l' we can ensure that B_1' and A' contain the same number of points of P. The resulting lines define $p_1 := l_1 \cap l_2$.

above l_1 or completely right of l_2. By a simple counting argument, any two small boxes not containing p_1 intersect. Any small box intersects any big box as a consequence of inequality (i); hence, applying Observation 4.1 we find p_2 satisfying the conditions of the Theorem.

For higher dimensions, we use hyperplanes instead of lines and we repeat the second step $d-1$ times (once in every direction except the first). ◀

A similar spitting idea works for weighted ε-nets of size 3: Let l_1 be a horizontal halving line and let l_2 be a vertical halving line. Let A and B be the areas above and below l_1 and let L and R be the areas left and right of l_2. The lines define four quadrants, where two opposite ones, say $L \cap A$ and $R \cap B$, both contain at least $\frac{n}{4}$ points of P. Define $p_1 := l_1 \cap l_2$. For every relevant box $\square$, assign $\square$ to the area $X \in \{A, B, L, R\}$ for which $|\square \cap X|$ is maximized. Put every box assigned to A and L into $\mathcal{A}$ and every box assigned to B and U into $\mathcal{B}$. Choosing the right values for $\varepsilon_1, \varepsilon_2$ and ε_3, we can apply Observation 4.1 to $\mathcal{A}$ and $\mathcal{B}$ to get the following:

▶ **Theorem 4.3.** *Let P be a set of n points in the plane. Let $0 < \varepsilon_1 \leq \varepsilon_2 \leq \varepsilon_3 < 1$ be arbitrary constants with (i) $\varepsilon_1 \geq \frac{3}{8}$, (ii) $\varepsilon_2 \geq \frac{1}{2}$, and (iii) $\varepsilon_1 + \varepsilon_3 \geq 1$. Then there exist three points p_1, p_2 and p_3 in $\mathbb{R}^2$ such that every axis-parallel box containing more than $\varepsilon_i n$ points of P contains at least i of the points p_1, p_2 and p_3.*

5 Conclusion

We have given bounds for weak weighted ε-nets of size 2 for convex sets and axis-parallel boxes. It remains an interesting question to find bounds for larger sizes. For axis-parallel boxes, we gave a construction for weighted ε-nets of size 3 in the plane. Unfortunately our construction does not generalize to higher dimensions. It is a natural question whether a similar statement in higher dimensions can be shown using a different construction.

References

1 N. Alon, I. Bárány, Z. Füredi, and D. J. Kleitman. Point Selections and Weak ε-Nets for Convex Hulls. *Combinatorics, Probability and Computing*, 1(3):189–200, 1992. URL: https://doi.org/10.1017/S0963548300000225.

2 B. Aronov, F. Aurenhammer, F. Hurtado, S. Langerman, D. Rappaport, C. Seara, and S. Smorodinsky. Small weak epsilon-nets. *Computational Geometry*, 42(5):455 – 462, 2009. URL: https://doi.org/10.1016/j.comgeo.2008.02.005.

3 Daniel Bertschinger and Patrick Schnider. Weighted epsilon-nets, 2020. arXiv:2002.08693.

4 B. Bukh and G. Nivasch. One-Sided Epsilon-Approximants. In *A Journey Through Discrete Mathematics: A Tribute to Jiří Matoušek*, pages 343–356. Springer, 2017. URL: https://doi.org/10.1007/978-3-319-44479-6_12.

5 S. Har-Peled. *Geometric Approximation Algorithms.* American Mathematical Society, 2011. URL: https://dlnext.acm.org/doi/10.5555/2031416.

6 S. Har-Peled and M. Sharir. Relative (p,ε)-Approximations in Geometry. *Discrete & Computational Geometry*, 45(3):462–496, 2011. URL: https://doi.org/10.1007/s00454-010-9248-1.

7 D. Haussler and E. Welzl. ε-nets and simplex range queries. *Discrete & Computational Geometry*, 2(2):127–151, 1987. URL: https://doi.org/10.1007/BF02187876.

8 J. Matoušek, E. Welzl, and L. Wernisch. Discrepancy and approximations for bounded VC-dimension. *Combinatorica*, 13(4):455–466, 1993. URL: https://doi.org/10.1007/BF01303517.

9 N. H. Mustafa and S. Ray. An optimal extension of the centerpoint theorem. *Computational Geometry*, 42(6):505 – 510, 2009. URL: https://doi.org/10.1016/j.comgeo.2007.10.004.

10 A. Pilz and P. Schnider. Extending the Centerpoint Theorem to Multiple Points. In *29th International Symposium on Algorithms and Computation (ISAAC 2018)*, volume 123, pages 53:1–53:13, 2018. URL: http://drops.dagstuhl.de/opus/volltexte/2018/10001.

11 R. Rado. A Theorem on General Measure. *Journal of the London Mathematical Society*, s1-21(4):291–300, 10 1946. URL: https://dx.doi.org/10.1112/jlms/s1-21.4.291.

12 N. Rubin. An Improved Bound for Weak Epsilon-Nets in the Plane. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, 2018. URL: https://doi.org/10.1109/FOCS.2018.00030.

13 V. Vapnik and A. Chervonenkis. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971. URL: https://doi.org/10.1137/1116025.