The Computational Complexity of Densest Region Detection

Shai Ben-David and **Nadav Eiron**
Department of Computer Science
Technion
Haifa 32000, Israel
{shai,nadav}@cs.technion.ac.il

Hans Ulrich Simon
Fakultät für Mathematik
Ruhr Universität Bochum
D-44780 Bochum, Germany
simon@lmi.ruhr-uni-bochum.de

March 5, 2001

Abstract

We investigate the computational complexity of the task of detecting dense regions of an unknown distribution from un-labeled samples of this distribution. We introduce a formal learning model for this task that uses a hypothesis class as its 'anti-overfitting' mechanism.

The learning task in our model can be reduced to a combinatorial optimization problem. We can show that for some constants, depending on the hypothesis class, these problems are $\mathcal{NP}$-hard to approximate to within these constant factors.

We go on and introduce a new criterion for the success of approximate optimization geometric problems. The new criterion requires that the algorithm competes with hypotheses only on the points that are separated by some margin μ from their boundaries.

Quite surprisingly, we discover that for each of the two hypothesis classes that we investigate, there is a 'critical value' of the margin parameter μ. For any value below the critical value the problems are $\mathcal{NP}$-hard to approximate, while, once this value is exceeded, the problems become poly-time solvable.

1 Introduction

Un-supervised learning is an important area of practical machine learning. Just the same, the computational learning theory literature has hardly addressed this issue. Part of this discrepancy may be due to the fact that there is no formal well defined model that captures the many different tasks that fall into this category. While the formation of such a comprehensive model may be a very difficult task, its absence should not deter the COLT community from researching models that capture restricted subareas of un-supervised learning. In this paper we investigate the computational complexity aspects of a formal model that addresses one specific task in this domain.

The model we discuss addresses the problem of locating the densest sub-domains of a distribution on the basis of seeing random samples generated by that distribution. This is, undoubtedly, one of the applicable tasks of un-supervised learning.

The scenario that we address is one in which the learner is supposed to infer information about an unknown distribution from a random sample it generates. An adequate model should therefore include some mechanism for avoiding over-fitting. That is, the model should impose some restrictions on the class of possible learner's outputs. The model we propose fixes a collection of domain subsets (a hypothesis class, if you wish) ahead of seeing the data. The task of the learner is to find a member of this class in which the average density of the example-generating distribution is maximized. For simplicity we restrict our attention to the case that the domain is the Euclidean space $\Re^n$. Density is defined relative to the Euclidean volume. By restricting our hypothesis classes to classes in which all the sets have the same volume, we can ignore the volume issue.

A model similar to ours was introduced by Ben-David and Lindenbaum [3]. In that paper a somewhat more general learning task is considered: given a threshold $r \in [0, 1]$, the learner is required to output the hypothesis in the class that best approximates the area on which the distribution has density above r. Ben-David et al. define a notion of a cost of a hypothesis, relative to a target distribution, and prove (ϵ, δ) type generalization bounds. As can be expected, the sample size needed for generalization depends on the VC-dimension of the underlying hypothesis class. We refer to that paper for a discussion of the relevance and potential applications of the model. However, [3] does not address the computational complexity of learning in this model.

Standard uniform convergence considerations imply that detecting a hypothesis (domain subset from the hypothesis class) with close-to-maximal density is essentially equivalent to detecting a hypothesis that approximates the maximal empirical density, with respect to the training data. We are therefore led to the following, purely combinatorial, problem:

> *Given a collection $\mathcal{H}$ of subsets of some domain set, on input – a finite subset P of the domain – output a set $h \in \mathcal{H}$ that maximizes $|P \cap h|$.*

We consider two hypothesis classes: the class of axis aligned hypercubes and the class of balls (both in $\Re^n$). For each of these classes we prove that there exists some $\gamma > 0$ (independent of the input sample size and dimensionality) such that, unless $\mathcal{P} = \mathcal{NP}$, no polynomial time algorithm can output, for every input sample, a hypothesis in the class that has agreement rate (on the input) within a factor of γ of the optimal hypothesis in the class.

On the other hand, we consider a relaxation of the common success criterion of optimization or approximation algorithms. Rather than requiring an approximation algorithm to achieve a fixed success ratio over all inputs (or over all inputs of the same size or dimensionality), we let the required approximation ratio depend on the structure of each specific input. Given a hypothesis class $\mathcal{H}$ of subsets of $\cup_n \Re^n$, and a parameter $\mu > 0$,

> an algorithm solves the μ-relaxed problem associated with $\mathcal{H}$, if, for every input sample, it outputs a member of $\mathcal{H}$ that contains as many sample points as any member of $\mathcal{H}$ can contain *with margin* $\geq \mu$ (where the margin of a point relative to a hypothesis is the radius of the largest ball around the point that is fully contained in the hypothesis).

In other words, such an algorithm is required to output a hypothesis with close-to-optimal performance on the input data, whenever this input sample allows a maximal intersection (with a member of $\mathcal{H}$) that achieves large enough margin for most of the points it contains. On the other hand, if for every element $h \in \mathcal{H}$ that achieves close-to-maximal-size intersection with the input a large percentage of the points in the intersection are close to h's boundaries, then an algorithm can settle for a relatively poor success ratio without violating the μ-relaxed criterion.

One appealing feature of this new performance measure is that it provides a rigorous success guarantee for agnostic learning that may be achieved by efficient algorithms for classes that can't have poly-time algorithms that succeed with respect to the common 'uniform' approximation ratio criterion. We shall show below that the class of balls provides such an example, and in a forthcoming paper [4] we show that the class of linear perceptrons is another such case.

This paper investigates the existence of poly-time algorithms that solve the μ-relaxed problem associated with a hypothesis class $\mathcal{H}$. Clearly, a relaxation becomes computationally easier as μ grows, and is hardest for $\mu = 0$, in which case it becomes the usual optimization problem (without relaxation). As mentioned above, we show that these optimization problems – finding the densest ball or the densest hypercube – are $\mathcal{NP}$-hard, to approximate (for other $\mathcal{NP}$-hardness results of this type see [9], [10]). We are interested in determining the values of μ at which the $\mathcal{NP}$-hardness of the relaxed problems breaks down.

Quite surprisingly, for each of the classes we investigate (axis-aligned hyper-cubes and balls), there exists a value μ_0 so that, on one hand, for every $\mu > \mu_0$, there exist efficient algorithms for the μ-relaxation, while on the other hand, for every $\mu < \mu_0$ the μ-relaxed problem is $\mathcal{NP}$-hard (and, in fact, even hard to approximate). A similar phenomenon holds also for the class of linear perceptrons [4].

The paper is organized as follows: Section 2 introduces the combinatorial optimization problems that we shall be considering, along with some basic background in hardness-of-approximation theory. Section 3 discusses the class of hypercubes and provides both the positive algorithmic result and the negative hardness result for this class. Next we discuss the class of balls. Section 4 contains the hardness result for this class while the following Section 5 provides efficient optimization algorithms for the μ-relaxation of the densest ball problem. Finally, in Section 6 we list several possible extensions of this work.

2 Definitions and Basic Results

In this section we introduce the combinatorial problems that we shall address in the paper. We then proceed to provide the basic definitions and tools that we shall use from the theory of approximation of combinatorial optimization problems. We end this section with a list of previously known hardness-of-approximation results that we shall employ in our work.

2.1 The Combinatorial Optimization Problems

We discuss combinatorial optimization problems of the following type:

The densest set problem for a class $\mathcal{H}$: Given a collection $\mathcal{H} = \cup_{n=1}^{\infty} \mathcal{H}_n$ of subsets, $\mathcal{H}_n \subseteq 2^{\Re^n}$, on input (n, P), where P is a finite multi-set of points in $\Re^n$, output a set $h \in \mathcal{H}_n$ so that h contains as many points from P as possible (accounting for their multiplicity in P).

We shall mainly be concerned with the following instantiations of the above problem:

Densest Axis-aligned Cube (DAC) Each class $\mathcal{H}_n$ consists of all cubes with side length equal 1 in $\Re^n$. That is, each member of $\mathcal{H}_n$ is of the form $\prod_{i=1}^{n} I_i$, where the I_i's are real intervals of the form $I_i = [a_i, a_i + 1]$.

Densest Open Ball (DOB) Each class $\mathcal{H}_n$ is the class of all open balls of radius 1 in $\Re^n$.

Densest Closed Ball (DCB) Each class $\mathcal{H}_n$ is the class of all closed balls of radius 1 in $\Re^n$.

For the sake of our proofs, we shall also address some other optimization problems, namely:

MAX-E2-SAT [1] Input is a collection C of 2-clauses over n Boolean variables. The problem is to find an assignment $a \in \{0,1\}^n$ satisfying as many 2-clauses of C as possible.

BSH Inputs are of the form (n, P_+, P_-), where $n \geq 1$, and P_+, P_- are multi-sets of points from $\Re^n$. A hyper-plane $H(w,t)$, where $w \in \Re^n$ and $t \in \Re$, *correctly classifies* $p \in P_+$ if $wp > t$, and it *correctly classifies* $p \in P_-$ if $wp < t$. The problem is to find the *Best Separating Hyper-plane* for P_+ and P_-, that is, a pair $(w,t) \in \Re^n \times \Re$ such that $H(w,t)$ correctly classifies as many points from $P_+ \cup P_-$ as possible.

DOH This is the densest set problem for the class of open hemispheres. That is, inputs are multi-sets P of points from S^n — the unit sphere in $\Re^n$ — and each class $\mathcal{H}_n$ is the class of all sets of the form $\{x : wx > 0\}$ for $w \in \Re^n$.

2.2 Basic Notions of Combinatorial Optimization

For each maximization problem Π and each input instance $\mathcal{I}$ for Π, $\mathrm{opt}_\Pi(\mathcal{I})$ denotes the maximum profit that can be realized by a legal solution for $\mathcal{I}$. Subscript Π is omitted when this does not cause confusion. The profit realized by an algorithm A on input instance $\mathcal{I}$ is denoted by $A(\mathcal{I})$. The profit associated with a legal solution σ for input instance $\mathcal{I}$ is denoted by $|\sigma|$. The quantity

$$\frac{\mathrm{opt}_\Pi(\mathcal{I}) - A(\mathcal{I})}{\mathrm{opt}_\Pi(\mathcal{I})} \tag{1}$$

is called the *relative loss of algorithm A on input instance $\mathcal{I}$*. Ideally, the relative loss is not much bigger than zero.

We generalize the definition above by allowing the performance of the algorithm to be measured relative to a function other than opt_Π. We denote the *V-relaxation of* Π by $\Pi[V]$. Analogously to (1) above, we define the *V-relative loss* of an algorithm A as

$$\frac{V(\mathcal{I}) - A(\mathcal{I})}{V(\mathcal{I})}. \tag{2}$$

An algorithm A is called *δ-approximation algorithm for* $\Pi[V]$, where $\delta \in \Re$, if its V-relative loss on $\mathcal{I}$ is at most δ for all input instances $\mathcal{I}$. Note that the original problem Π is the same as the opt_Π-relaxation of Π.

Let $\Pi[V]$ and $\Pi'[V']$ be two (potentially relaxed) maximization problems. A *polynomial reduction* from $\Pi[V]$ to $\Pi'[V']$, consists of two functions:

Input Transformation a polynomial time computable mapping $\mathcal{I} \mapsto \mathcal{I}'$, which transforms an input instance $\mathcal{I}$ of Π into an input instance $\mathcal{I}'$ of Π'

Solution Transformation a polynomial time computable mapping $(\mathcal{I}, \sigma') \mapsto \sigma$, which transforms $(\mathcal{I}, \sigma')$, where $\mathcal{I}$ is an input instance of Π and σ' is a legal solution for $\mathcal{I}'$, into a legal solution σ for $\mathcal{I}$

[1] "E2" stands for "Exactly 2 literals per clause".

We write $\Pi[V] \leq_{\rm pol} \Pi'[V]$ to indicate that there exists a polynomial reduction of $\Pi[V]$ to $\Pi'[V]$. Given a polynomial time algorithm A' which finds a legal solution σ' for each given input instance $\mathcal{I}'$ of Π' and a polynomial reduction from Π to Π', we obtain the following polynomial time algorithm A for Π:

1. Compute $\mathcal{I}'$ from $\mathcal{I}$.
2. Compute a legal solution σ' for $\mathcal{I}'$ using A'.
3. Compute a legal solution σ for $\mathcal{I}$ from $(\mathcal{I}, \sigma')$.

We refer to A as the algorithm *induced* by A' and the reduction.

In general, a polynomial reduction is not approximation-preserving. Even if A' is a δ-approximation algorithm for $\Pi'[V']$, there is in general no upper bound on the V-relative loss of the induced algorithm A. In this paper, we shall use special reductions which obviously are approximation-preserving:

Definition 2.1 *Assume that* $\Pi[V] \leq_{pol} \Pi'[V']$. *We say that there is a* loss-preserving *reduction of* $\Pi[V]$ *to* $\Pi'[V]$, *written as* $\Pi[V] \leq_{pol}^{lp} \Pi'[V']$, *if there exists a polynomial reduction that satisfies the following conditions:*

1. *The input transformation maps* $\mathcal{I}$ *to* $\mathcal{I}'$ *such that* $V'(\mathcal{I}') \geq V(\mathcal{I})$.
2. *The solution transformation maps* $(\mathcal{I}, \sigma')$ *to* σ *such that* $|\sigma| \geq |\sigma'|$.

The following result motivates the name "loss-preserving".

Lemma 2.2 *If* $\Pi[V] \leq_{pol}^{lp} \Pi'[V']$ *and there is no polynomial time* δ*-approximation algorithm for* $\Pi[V]$, *then there is no polynomial time* δ*-approximation algorithm for* $\Pi'[V']$.

Proof: Assume for sake of contradiction that A' is a polynomial time δ-approximation algorithm for $\Pi'[V']$. Let A be the algorithm induced by A' and and a loss-preserving reduction of $\Pi[V]$ to $\Pi'[V]$. The definition of loss-preserving reductions implies that $V'(\mathcal{I}') \geq V(\mathcal{I})$ and $A(\mathcal{I}) \geq A'(\mathcal{I}')$. Thus,

$$\frac{V(\mathcal{I}) - A(\mathcal{I})}{V(\mathcal{I})} = 1 - \frac{A(\mathcal{I})}{V(\mathcal{I})} \leq 1 - \frac{A'(\mathcal{I}')}{V'(\mathcal{I}')} = \frac{V'(\mathcal{I}') - A'(\mathcal{I}')}{V'(\mathcal{I}')} \leq \delta.$$

We arrived at a contradiction. □

2.3 Relaxed Densest Set Problems

As mentioned in the introduction, we shall mainly discuss a new notion of relaxation for densest set problems. The idea behind this new notion is that the required approximation rate varies with the structure of the input sample. When there exist optimal solutions that are 'stable', in the sense that minor variations to these solutions will not affect their cost, then we require a high approximation ratio. On the other hand, when all optimal solutions are 'unstable' then we settle for lower approximation ratios. This idea is formalized by comparing the profit of the approximation algorithm not to the cost of the optimal solution (i.e., the number of input points included in the optimal solution), but rather to the number of points from the input that the optimal solution contains with some margin μ.

Before we proceed, let us fix some notation. Let $n \geq 1$, $w, z \in \Re^n$, $t \in \Re$, and $r \in \Re^+$. $H(w,t) = \{x \in \Re^n : wx = t\}$ denotes the hyper-plane induced by w and t. $H_+(w,t) = \{x \in \Re^n : wx > t\}$ and $H_-(w,t) = \{x \in \Re^n : wx < t\}$ denote the corresponding positive and negative open half-space, respectively. $B(z,r) = \{x \in \Re^n : \|z - x\| < r\}$ denotes the open ball of radius r around center z. $\bar{B}(z,r) = \{x \in \Re^n : \|z - x\| \leq r\}$ denotes the corresponding closed ball. z^n denotes the all-zeroes vector in $\Re^n$ (the origin). $B^n = B(z^n, 1)$ is our short notation for the open unit ball, $\bar{B}^n = \bar{B}(z^n, 1)$ denotes the closed unit ball, and $S^n = \{x \in \Re^n : \|x\| = 1\}$ denotes the unit sphere in $\Re^n$.

Definition 2.3 *Let $\mathcal{H} = \cup_{n \geq 1} \mathcal{H}_n$, where $\mathcal{H}_n \subseteq 2^{\Re^n}$, be a hypothesis class, and let $\mu > 0$ be a positive real.*

- *For each $h \in \mathcal{H}_n$, let h^μ be the set of points that are* included in h with a margin μ, *i.e.,*

$$h^\mu \stackrel{\triangle}{=} \{x \in \Re^n \ : \ \bar{B}(x, \mu) \subseteq h\}.$$

- *Given a finite multiset $P \subset \Re^n$, let*

$$V_\mu(n, P) \stackrel{\triangle}{=} \max_{h \in \mathcal{H}_n} |P \cap h^\mu|.$$

  In other words, $V_\mu(n,P)$ denotes the maximum number of points in P (accounting for their multiplicity) that can be included in a hypothesis from $\mathcal{H}_n$ with margin μ.

- *The* μ-relaxed densest set problem *for $\mathcal{H}$ is defined as the V_μ-relaxation of the densest set problem for $\mathcal{H}$.*

We use $\Pi[\mu]$ to denote the μ-relaxation $\Pi[V_\mu]$ of the problem Π.

2.4 Some Known Hardness-of-Approximation Results

We shall base our hardness reductions on two known results.

Theorem 2.4 [Håstad, [8]] *Assuming $\mathcal{P} \neq \mathcal{NP}$, for any $\delta < 1/22$, there is no polynomial time δ-approximation algorithm for MAX-E2-SAT.*

Theorem 2.5 [Ben-David, Eiron and Long, [2]] *Assuming $\mathcal{P} \neq \mathcal{NP}$, for any $\delta < 3/418$, there is no polynomial time δ-approximation algorithm for BSH.*

Claim 2.6 *BSH$\leq^{lp}_{pol}$DOH.*

Proof: By adding a coordinate one can translate hyper-planes to *homogeneous* hyper-planes (i.e., hyper-planes that pass through the origin). To get from the homogeneous hyper-planes separating problem to the densest hemisphere problem one applies the standard scaling and reflection tricks. □

Corollary 2.7 *Assuming $\mathcal{P} \neq \mathcal{NP}$, there is no polynomial time δ-approximation algorithm for DOH, for any $\delta < 3/418$.*

3 The Relaxed Densest Cube Problem

We present a (rather simple) algorithm, Algorithm 3.1 below, which solves the 1/4-relaxation of the DAC problem in polynomial time. We complement this result by showing that the μ-relaxation of DAC is $\mathcal{NP}$-hard (and, in fact, even $\mathcal{NP}$-hard to approximate) for each $\mu < 1/4$. Thus (despite its simplicity), Algorithm 3.1 already solves the hardest relaxation of DAC that can be solved by any polynomial time algorithm (unless $\mathcal{P} = \mathcal{NP}$).

An axis aligned cube with edge length u will be briefly called *u-cube* in what follows. Let W be a 1-cube. Note that the points which are contained in W with margin $0 \le \mu \le 1/2$ are contained in a concentric $(1-2\mu)$-sub-cube of W. (Figure 1 illustrates this observation.)

Recall that the input to DAC has the form (n, P), where P is a finite multi-set of points from $\Re^n$. The legal outputs are the n-dimensional 1-cubes. Recall furthermore that $V_\mu(n, P)$ denotes the maximal number of points from P (accounting for their multiplicity) that can be included in a 1-cube with margin μ. According to our observation above, $V_{1/4}(n, P)$ coincides with the maximal number of points from P that can be included in a 1/2-cube. In order to solve the 1/4-relaxation of DAC, one has to output a 1-cube containing at least $V_{1/4}(n, P)$ points. Here is the algorithm which achieves this goal:

Algorithm 3.1:

1. For each point p in P do:
   (a) Let $W[p]$ be the 1-cube with center p.
   (b) Compute the "gain" $G[p] \triangleq |W[p] \cap P|$ associated with each $p \in P$.
2. Choose the point $p_{\max}$ such as to maximize $G[p]$ and output $W[p_{\max}]$.

Theorem 3.2 *Algorithm 3.1 solves the* $1/4$*-relaxation of DAC and runs in polynomial time.*

Proof: Let W_* be the 1/2-cube which contains $V_{1/4}(n, P)$ points. Let p_* be a point from $W_* \cap P$ (arbitrarily chosen). Obviously $W[p_*]$ (the 1-cube with center p_*) contains W_* as sub-cube. (Figure 1 illustrates this observation.) Therefore,

$$G(p_{\max}) = |W[p_{\max}] \cap P| \ge |W[p_*] \cap P| \ge |W_* \cap P| = V_{1/4}(n, P).$$

It follows that Algorithm 3.1 solves the 1/4-relaxation of DAC. It clearly runs in polynomial time. $\square$

The next result shows that the 1/4-relaxation of DAC is the hardest relaxation which can be solved in polynomial time.

Theorem 3.3 *For every* $0 \le \mu < 1/4$, *MAX-E2-SAT*$\le_{pol}^{lp}$*DAC*$[\mu]$.

Proof: Let $0 \le \mu < 1/4$ be fixed. Remember that we can view 1-cubes, containing a set of points with margin μ, as $(1-2\mu)$-cubes containing the same set of points. The point in $\Re^n$ whose i-th coordinate equals $1-2\mu$ and whose other coordinates equal zero is denoted by p_i^μ in what follows.

We will define a loss-preserving reduction from MAX-E2-SAT to the μ-relaxation of DAC. (Compare with Definition 2.1.) The following constructions are illustrated in Figure 2.

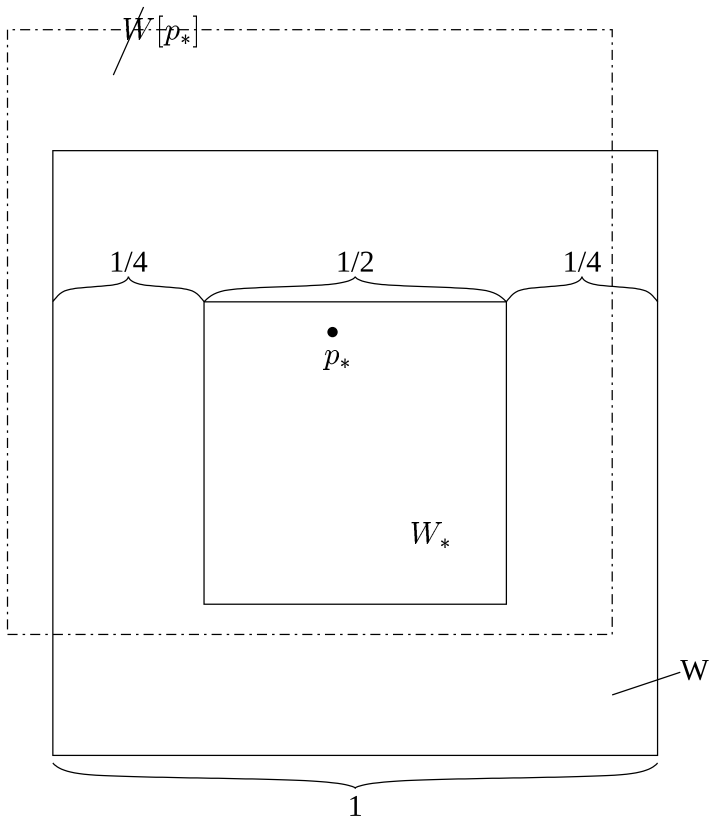


Figure 1: A 1-cube W, its concentric $1/2$-sub-cube W_*, a point $p_* \in W_*$, and the 1-cube $W[p_*]$ centered at p_*.

First, we define a mapping ϕ from input instances of MAX-E2-SAT (over n variables) to finite multi-sets in $\Re^n$. Let $v_1, \ldots, v_n$ be the variables that appear in the propositional formulas, and let $\neg v_1, \ldots, \neg v_n$ denote their negations, respectively. Let h be the function which maps v_i to p_i^μ and $\neg v_i$ to $-p_i^\mu$. Given a 2-clause $l_1 \vee l_2$, define the point triplet

$$\phi(l_1 \vee l_2) \stackrel{\triangle}{=} \{h(l_1) + h(l_2), -h(l_1) + h(l_2), h(l_1) - h(l_2)\}.$$

Informally, we associate with each 2-clause c three points in the plane spanned by the coordinates that correspond to the variables of c. These three points represent the three satisfying assignments for c. Finally, we extend the definition of ϕ from 2-clauses to collections of 2-clauses by setting

$$\phi(C) \stackrel{\triangle}{=} \bigcup_{c \in C} \phi(c),$$

where the union of sets should be interpreted as multi-set. Obviously, $\phi(C)$ can be constructed from C in polynomial time.

Let $a = (a_1, \ldots, a_n) \in \{0,1\}^n$ be an assignment to $(v_1, \ldots, v_n)$ which satisfies the maximal number, say s_*, of 2-clauses from C. We have to show that there exists a $(1 - 2\mu)$-cube which contains at least s_* points of the multi-set $\phi(C)$. Setting $I[0] = [-(1-2\mu), 0]$ and $I[1] = [0, 1-2\mu]$, the $(1-2\mu)$-cube $W_a = \times_{i=1}^n I[a_i]$ serves this purpose. More precisely, W_a contains one of the three points of $\phi(c)$ iff a satisfies c. Thus, W_a contains s_* points of $\phi(C)$.

Let now $W = \times_{i=1}^n I_i$ be a 1-cube which contains s points from $\phi(C)$. Note that each cube which contains at least two points of the point triplet $\phi(c)$ must have side length at least $2(1-2\mu) > 1$. Thus, W contains at most one point of each triplet. It follows that there are s "designated" 2-clauses in C with the property that one point of the associated point triplet belongs to W. Setting $a_i = 1$ if interval I_i contains $1-2\mu$ and $a_i = 0$ otherwise, we obtain an assignment $a = (a_1, \ldots, a_n)$ to $(v_1, \ldots, v_n)$ which satisfies precisely the s designated 2-clauses. Clearly, a can be computed from C and W in polynomial time. This completes the loss-preserving reduction from MAX-E2-SAT to DAC$[\mu]$. □

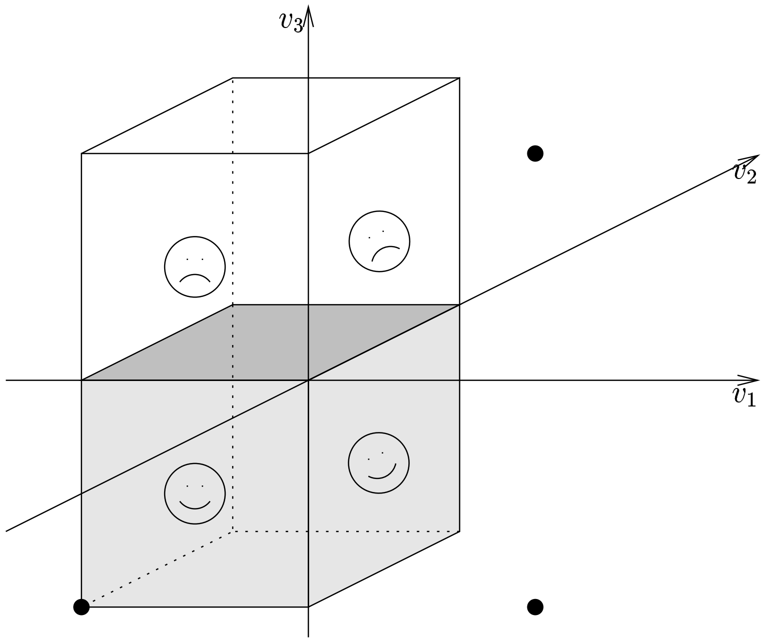


Figure 2: The point triplet associated with the 2-clause $v_1 \vee \neg v_3$ in the $\{v_1, v_2, v_3\}$-space, the cube corresponding to the satisfying assignment $(v_1, v_2, v_3) = (0, 1, 0)$, and (at top of it) the cube corresponding to the falsifying assignment $(v_1, v_2, v_3) = (0, 1, 1)$

Corollary 3.4 *For every $0 \le \mu < 1/4$ and for every $0 \le \delta < 1/22$, there is no polynomial time δ-approximation algorithm for the μ-relaxation of DAC (unless $\mathcal{P} = \mathcal{NP}$).*

4 Hardness of the Densest Ball Problem

In this section we prove a hardness-of-approximation result for the DOB problem (Theorem 4.2 below). We refer to $H_+(w, 0)$ as an *open hemisphere* because we use the hyper-plane $H(w, 0)$ as a separator of the unit sphere S^n into two hemispheres. We may assume that $\|w\| = 1$ because for all $\lambda > 0$, $H_+(w, 0) = H_+(\lambda w, 0)$.

Lemma 4.1 *DOH$\le_{pol}^{lp}$DOB.*

Proof: Let $\mathcal{I} = (n, P)$ be a given input to DOH, where P is a multi-set of points in S^n. We choose the trivial input transformation $\mathcal{I} \mapsto \mathcal{I}$, i.e., $\mathcal{I} = (n, P)$ is also considered as input to DOB.

Let $C(w, P)$ be the multi-set of points from P that also belong to $H_+(w, 0)$, and let $C'(z, P)$ be the multi-set of points from P that also belong to $B(z, 1)$. The reduction from DOH to DOB is now accomplished by proving the following statements:

$$\forall w \in S^n, \exists z \in \Re^n : \quad C(w, P) \subseteq B(z, 1) \tag{3}$$

$$\forall z \in \Re^n : \quad C'(z, P) \subseteq H_+(z, 0) \tag{4}$$

These statements certainly imply that there is a loss-preserving reduction from DOH to DOB. (Compare with Definition 2.1.)

To prove statement (3), we set $\mu = \min_{p \in C(w,P)} |w \cdot p|$. This implies that $w \cdot q \geq \mu > 0$ for all $q \in C(w, P)$. We claim that $z = \mu w$ is an appropriate choice for z, i.e., each $q \in C(w, P)$ also belongs to $B(z, 1)$. Using $w \cdot w = q \cdot q = 1$, this claim is evident from the following calculation:

$$\begin{aligned} \|z - q\|^2 &= (z - q) \cdot (z - q) \\ &= z \cdot z - 2z \cdot q + q \cdot q \\ &= \mu^2 w \cdot w - 2\mu w \cdot q + q \cdot q \\ &= \mu^2 - 2\mu w \cdot q + 1 \\ &\leq \mu^2 - 2\mu^2 + 1 \\ &= 1 - \mu^2 \\ &< 1 \end{aligned}$$

In order to prove statement (4), we have to show that each $q \in C'(z, P)$ satisfies $z \cdot q > 0$. To this end, note first that $q \in C'(z, P)$ implies $q \cdot q = 1$ and

$$1 > \|z - q\|^2 = z \cdot z - 2z \cdot q + q \cdot q = z \cdot z - 2z \cdot q + 1 \geq -2z \cdot q + 1.$$

Clearly, this implies that $z \cdot q > 0$. □

Applying Corollary 2.7 we readily get

Theorem 4.2 *Assuming $\mathcal{P} \neq \mathcal{NP}$, there is no polynomial time δ-approximation algorithm for DOB, for any $\delta < 3/418$.*

As shown in [4], a similar result holds for the Densest Closed Ball problem.

Theorem 4.3 (Ben-David and Simon, [4]) *Assuming $\mathcal{P} \neq \mathcal{NP}$, there is no polynomial time δ-approximation algorithm for DCB, for any $\delta < 1/198$.*

5 Computation of Dense Balls

We know from Section 4 that it is an NP-hard problem to find an (approximately) densest (open or closed) ball for a given multi-set of points in $\Re^n$. In this section, we show that, for each constant $0 < \mu \leq 1$, the μ-relaxation of this problem can be solved optimally in polynomial time. For the sake of exposition, we restrict the following discussion to closed balls. Also, for brevity, we use *r-ball* to refer to a ball with radius r.

Let $\bar{B}$ be a 1-ball. Note that the points which are contained in $\bar{B}$ with margin μ form the concentric $(1-\mu)$-sub-ball of $\bar{B}$. It follows that, an algorithm that solves the μ-relaxation of DCB on input (n, P) must output the center of a 1-ball $\bar{B}$ such that $|\bar{B} \cap P| \geq |\bar{B}_* \cap P|$ for every $(1-\mu)$-ball $\bar{B}_*$. A simple scaling argument shows that instead of balls with radius 1 and $1-\mu$, respectively, we may as well consider balls with radius $\frac{1}{1-\mu}$ and 1, respectively. The goal is therefore to design a family of algorithms which can successfully compete against the densest closed 1-ball by means of a closed ball of a radius slightly exceeding 1. This general idea is captured by the following definitions.

Assume that $R(k, n)$ is a function which maps each pair (k, n) to a positive real $R(k, n) \in \Re^+$. R is called *admissible* if $\lim_{k\to\infty} \lim_{n\to\infty} R(k, n) = 0$.

A family $(A_k)_{k\geq 1}$ of polynomial time algorithms is called *R-successful* for DCB if, on input (n, P), A_k outputs the center of a $(1 + R(k, n))$-ball $\bar{B}$ such that $|\bar{B} \cap P| \geq |\bar{B}_* \cap P|$ for every 1-ball $\bar{B}_*$.

Lemma 5.1 *If there exists a family of polynomial time algorithms which is R-successful for an admissible function R, then the μ-relaxation of DCB can be solved in polynomial time for each $\mu > 0$.*

Proof: Let (n, P) be the input to DCB. Choose k such that $R(k, n) \leq \frac{\mu}{1-\mu}$. This is always possible, for large enough n, since R is admissible. Define scaling factor

$$\lambda \triangleq \frac{1}{1-\mu} = 1 + \frac{\mu}{1-\mu}.$$

Note that $1 + R(k, n) \leq \lambda$ and $1/\lambda = 1 - \mu$. Apply the algorithm A_k (the k-th member of the R-successful family) to input $(n, \lambda \cdot P)$, where $\lambda \cdot P \triangleq \{\lambda \cdot p | \, p \in P\}$. If A_k outputs center z, then output center $\frac{1}{\lambda} z$. Since A_k belongs to an R-successful family, $\bar{B}(z, 1 + R(k, n))$ does not contain less points from $\lambda \cdot P$ than any 1-ball. We can make the same claim à-fortiori for $\bar{B}(z, \lambda)$. It follows that $\bar{B}(z/\lambda, 1)$ does not contain less points from P than any $(1-\mu)$-ball. $\square$

The main result of this section is:

Theorem 5.2 *For each $\mu > 0$, the μ-relaxation of DCB can be solved in polynomial time.*

According to Lemma 5.1, the theorem is obtained once we have presented an R-successful family of polynomial time algorithms for an admissible function R. To this end, we proceed as follows. In Subsection 5.1, we present a generic family of algorithms for DCB containing some programmable parameters. The appropriate setting of these parameters requires some geometric insights which are provided in Subsection 5.2. Three concrete (families of) algorithms for DCB are analyzed in Subsections 5.3 and 5.4: the Center-of-Gravity algorithm, the Smallest-Ball algorithm, and the Equal-Distance algorithm. The first algorithm is R_1-successful for $R_1(k, n) = \sqrt{1/k}$. The other two algorithms are both R_2-successful for $R_2(k, n) = \sqrt{\frac{n-k+1}{kn}}$. R_1 and R_2 are both admissible. Furthermore, $R_2(k, n) \leq \sqrt{1/k}$ with equality when n approaches infinity. According to Lemma 5.1, the members of the R_1- and R_2-successful families solve the μ-relaxations of DCB. For a given μ, it is sufficient to use an algorithm A_k such that $\sqrt{1/k} \leq \frac{\mu}{1-\mu}$. E.g., $k = \lceil 1/\mu^2 \rceil$ is a possible choice.

5.1 A Generic Algorithm for the Densest Ball Problem

Let Z denote a function which maps finite multi-sets Q of points from $\Re^n$ to points in $\Re^n$. Let R denote a function which maps pairs (k, n) to positive reals. Here is the high level description of our generic algorithm for DCB:

Algorithm 5.3: Let Z, R, k be fixed. Algorithm $A_k^{Z,R}$ proceeds as follows on input (n, P):

1. For each multi-set Q containing at most k points from P (possibly with repetitions), compute two quantities associated with Q:
   - the "candidate center" $Z(Q) \in \Re^n$,
   - the "gain" $G(Z(Q)) \triangleq |P \cap \bar{B}(Z(Q), 1 + R(k, n))|$.
2. Choose $Z_{\max}$ from the candidate centers such as to maximize $G(Z)$ and output $Z_{\max}$.

Before we present some concrete choices for the functions Z, R, we briefly mention the following potential simplifications of the generic algorithm:

Fixed Size Multisets (FSM) If the generic algorithm is run in *FSM-mode*, then we check only multi-sets Q of fixed size k.

No Repetitions (NR) If the generic algorithm is run in *NR-mode*, then we consider sets Q instead of multi-sets.

Here are three concrete choices for the Z- and the R-function, respectively, which will be analyzed in the course of the next subsections:

Center-of-Gravity (A_k^{CG}) uses the following functions in the roles of Z and R, respectively:

$$Z_{\mathrm{CG}}(Q) \triangleq \frac{1}{|Q|} \cdot \sum_{q \in Q} q \tag{5}$$

$$R_1(k, n) \triangleq \sqrt{1/k} \tag{6}$$

Smallest-Ball Let $Z_{\mathrm{SB}}(Q)$ be the center of the smallest ball containing Q and

$$R_2(k, n) \triangleq \sqrt{\frac{n - k + 1}{kn}} \tag{7}$$

(A_k^{SB}) uses Z_{SB} and R_2 in the roles of Z and R, respectively.

Equal-Distance Let $Z_{\mathrm{ED}}(Q)$ be the point which has the same Euclidean distance to all $q \in Q$ and belongs to the same affine sub-space as Q. (A_k^{ED}) uses Z_{ED} and R_2 in the roles of Z and R, respectively.

A_k^{CG} is run in FSM-mode. A_k^{SB} and A_k^{ED} are run in NR-mode.

Here are some brief comments on the computational complexity of these algorithms. Note first that it is not necessary to evaluate the square root. E.g., in order to check whether a point x is included in a ball of radius $\sqrt{1/k}$ around a center z, one checks whether the square of the Euclidean distance between x and z (which can be expressed as a scalar product) is bounded by $1/k$. Note second that all three algorithms perform an exhaustive search through $O(m^k)$ candidate (multi-)sets. For each fixed (multi-)set, there is a polynomial time bound. A_k^{CG} is the simplest algorithm, because each center of gravity is found in linear time. In order to find the center of the smallest ball that contains a given set Q, one has to solve a quadratic programming problem subject to linear constraints. In order to find $A_k^{\mathrm{ED}}(Q)$, it is sufficient to solve a system of linear equations (e.g., by Gaussian elimination) because $A_k^{\mathrm{ED}}(Q)$ can be expressed as the intersection of hyper-planes (after a coordinate transformation which maps the points in the appropriate sub-space). Thus, A_k^{SB} requires more computational resources than A_k^{ED}, which, in turn, requires more computational resources than A_k^{CG}.

All three algorithms share a common property: they are "compatible with translation and scaling". More precisely, let

$$z_0 + P \stackrel{\triangle}{=} \{z_0 + p \mid p \in P\} \text{ and } \lambda \cdot P \stackrel{\triangle}{=} \{\lambda \cdot p \mid p \in P\}$$

for each $z_0 \in \Re^n$ and $\lambda > 0$. We say that Z is *compatible with translation* if equation

$$Z(n, z_0 + P) = z_0 + Z(n, P)$$

is valid for each choice of n, P, z_0. We say that Z is *compatible with scaling* if equation

$$Z(n, \lambda \cdot P) = \lambda \cdot Z(n, P)$$

is valid for each choice of n, P, λ. The following result is fairly obvious:

Lemma 5.4 *The functions Z_{CG}, Z_{SB}, and Z_{ED} are compatible with translation and scaling.*

Algorithms which are compatible with translation and scaling exhibit a nice feature. Their performance analysis can be restricted w.l.o.g. to "normalized inputs". More precisely, we say that an input (n, P) for DCB is *normalized* if the following holds:

- The closed unit ball $\bar{B}^n$ is a densest closed 1-ball on input (n, P).
- Each closed ball of radius less than 1 contains less points from P than $\bar{B}^n$.

In other words, the closed unit ball is the (unique) smallest densest ball on input (n, P). The following result is obvious:

Lemma 5.5 *Let (A_k) be a family of algorithms for DCB which is compatible with translation and scaling. If (A_k) is R-successful on each normalized input, then (A_k) is R-successful (on each input).*

5.2 Full and Partial Spanning Sets for Smallest Balls

Let B be an open n-dimensional ball and $\bar{B}$ the corresponding closed ball. The center of B is denoted as z_B. The boundary of $\bar{B}$, called *B-sphere* hereafter, is denoted as S_B. Each hyper-plane H partitions $\Re^n$ into the three sets H_+, H_- (the open half-spaces induced by H) and H itself. Each hyper-plane H which passes through z_B cuts the B-sphere S_B into two *open B-hemispheres*, namely $S_B \cap H_+$ and $S_B \cap H_-$. Let $P \subseteq \Re^n$. The following lemma presents necessary and sufficient conditions for $\bar{B}$ being the smallest ball containing P.

Lemma 5.6 *Assume that P is contained in $\bar{B}$. The following statements are equivalent:*

A1 *$\bar{B}$ is the smallest ball containing P.*

A2 *The points in $P \cap S_B$ are not contained in any open B-hemisphere.*

A3 *The convex hull K of $P \cap S_B$ contains z_B.*

Proof: Let $\neg A1, \neg A2, \neg A3$ denote the negations of the statements $A1, A2, A3$, respectively. We will prove the implications $\neg A3 \Rightarrow \neg A2$, $\neg A2 \Rightarrow \neg A1$, and $A3 \Rightarrow A1$.

Let us start with the implication $\neg A3 \Rightarrow \neg A2$. Assume that K does not contain z_B. W.l.o.g. $K \neq \emptyset$ (because otherwise $P \cap S_B = \emptyset$ and $\neg A2$ trivially holds). The following reasoning is illustrated in Figure 3. Let p be the point in K with a minimal distance $d > 0$ to z_B, L_p the line segment from z_B to p, and H_p the hyper-plane through p that is orthogonal to L_p. The fact that K is a convex polyhedron and the choice of p imply that H_p does not intersect the interior of K. If H'_p denotes the hyper-plane that is parallel to H_p and passes through z_B (obtained by a parallel shift of H_p along L_p), then K is totally contained in one of the open half-spaces induced by H'_p. The points in $P \cap S_B$ are therefore contained in one of the open B-hemispheres induced by H'_p.

We proceed with the proof for the implication $\neg A2 \Rightarrow \neg A1$. The following reasoning is illustrated in Figure 3. Assume that H is a hyper-plane through z_B such that the points in $P \cap S_B$ are contained in one of the open B-hemispheres induced by H. Let R be the ray that starts in z_B and is directed orthogonally away from H towards the hemisphere containing $P \cap S_B$. Let B_ϵ be the open ball obtained from B by performing an ϵ-shift of center z_B along R. It follows that there exists an $\epsilon > 0$ such that B_ϵ contains P. Since B_ϵ is open, it can be shrunken and still contain P. It follows that $\bar{B}$ is not the smallest ball containing P.

We finally show the implication $A3 \Rightarrow A1$. Assume that $z_B \in K$. The following reasoning is illustrated in Figure 4. Let B' be a ball with center $z_{B'}$ such that $z_{B'} \neq z_B$ and $P \subseteq \bar{B}'$. We have to show that the radius r' of B' is greater than the radius r of B. Let L be the line segment from z_B to $z_{B'}$. Let H be the hyper-plane through z_B which is orthogonal to L, H_+ the open half-space containing $z_{B'}$ and H_- the other open half-space. Since $z_B \in K$, H_- must contain at least one point p of $P \cap S_B$. When we move a point z along L from z_B to $z_{B'}$, its distance to p strictly increases. Since the distance between z_B and p coincides with r and the distance between $z_{B'}$ and p is a lower bound on r', we get $r' > r$. □

Let $Q \subset \Re^n$ be a set of points in general position. It follows that $|Q| \leq n + 1$. The convex hull of Q is then called the *simplex induced by Q* and denoted as $\mathcal{S}(Q)$. Occasionally, we will blur the distinction between Q and the induced simplex and use the notation Q for both objects. Remember that each polyhedron in $\Re^n$ can be partitioned into simplexes (which corresponds to a triangulation of a polygon in the plane). The following definition relates simplexes to balls:

Definition 5.7 *Let $\bar{B}$ be a ball in $\Re^n$ and $Q \subset \Re^n$ be a set of points in general position. Q is called* spanning set *for $\bar{B}$ if $Q \subset S_B$ and $z_B \in \mathcal{S}(Q)$. A spanning set for an n-dimensional ball is called* degenerated *if it contains at most n points.*

The following results are (more or less) immediate consequences of Lemma 5.6.

Corollary 5.8 *If Q is a spanning set for $\bar{B}$, then $Z_{SB}(Q) = Z_{ED}(Q) = z_B$.*

Figure 5 illustrates Corollary 5.8.

Corollary 5.9 *Let $\bar{B}$ be the smallest n-dimensional ball containing a finite set P of points from $\Re^n$. Then $P \cap S_B$ contains a spanning set Q for $\bar{B}$.*

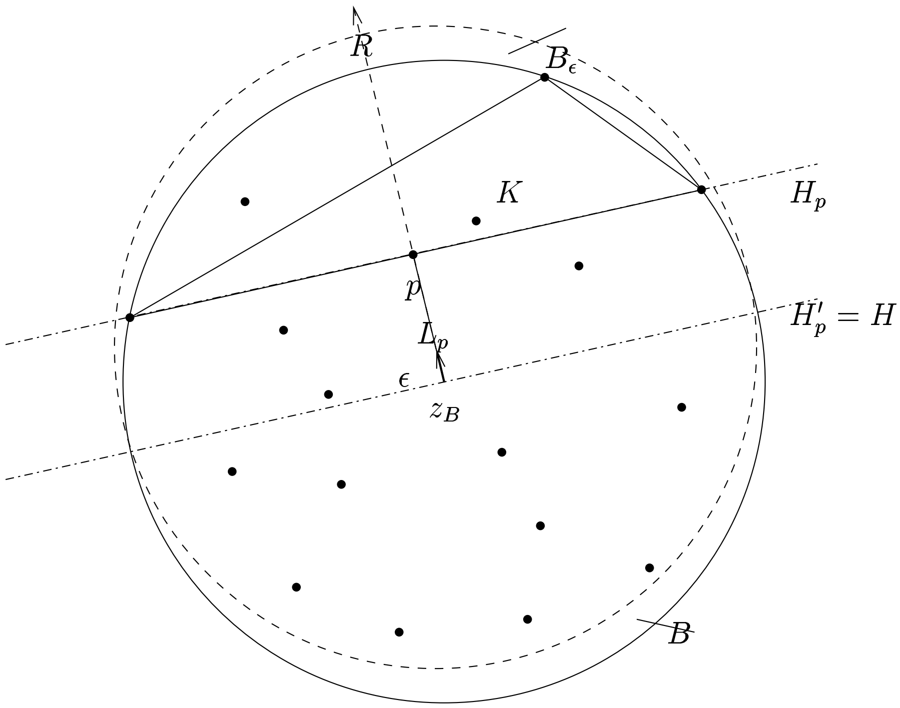


Figure 3: A non-smallest ball containing a given set of points.

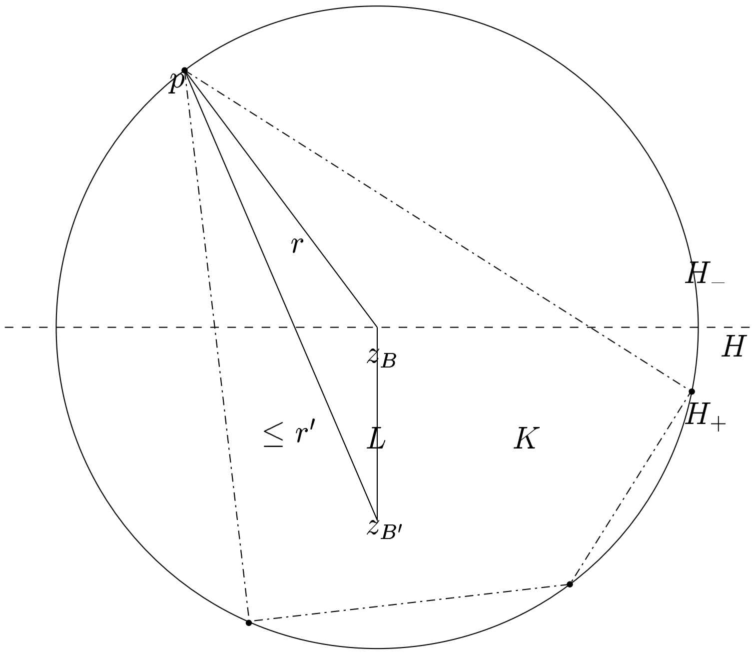
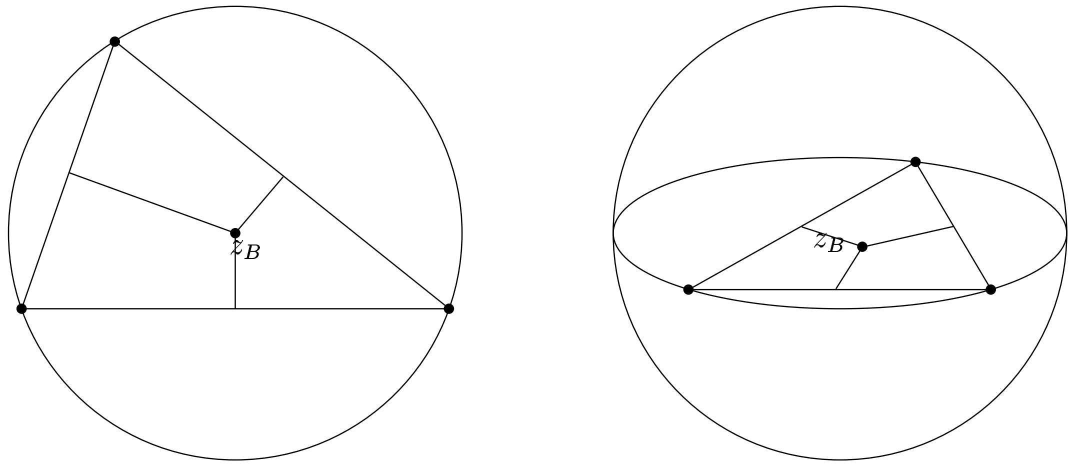


Figure 4: A smallest ball containing a given set of points.



(a) (b)

Figure 5: (a) A 2-dimensional example: a spanning set Q of size 3 for a 2-dimensional ball $\bar{B}$. (b) A 3-dimensional example: a degenerated spanning set Q of size 3 for a 3-dimensional ball $\bar{B}$. In both cases, $z_B = Z_{\mathrm{SB}}(Q) = Z_{\mathrm{ED}}(Q)$.

Proof: According to Lemma 5.6, the convex hull K of $P \cap S_B$ contains z_B. Form a simplicial decomposition of the polyhedron K. One of the simplexes in this decomposition, say $\mathcal{S}$, must contain z_B. It follows that the vertex set Q of $\mathcal{S}$ is a spanning set for $\bar{B}$. □

Let us briefly explain how we will use the concept of spanning sets. Remember that we can restrict ourselves to a normalized input (n, P), whose smallest densest ball is the unit ball $\bar{B}^n$. According to Corollary 5.9, $P \cap S^n$ contains a spanning set Q^n for $\bar{B}^n$. Remember that $|Q^n| \leq n+1$. We want to argue that the generic algorithm $A_k^{Z,R}$ computes at least one candidate center $Z(Q)$ that is close to the origin z^n. Note that $Z_{\mathrm{SB}}(Q^n) = Z_{\mathrm{ED}}(Q^n) = z^n$. Thus, the Smallest-Ball and the Equal-Distance algorithms would both find the optimal center if they inspected the full spanning set Q^n. However, the generic algorithm inspects only (multi-)sets Q of size at most k, and k is much smaller than n in general. The hope is that Q^n contains a small subset Q such that $Z(Q)$ comes already close to the origin. This motivates the following definition.

Definition 5.10 *Let R be a function which maps a pair (k, n) to a positive real $R(k, n) \in \Re^+$. We say that function Z is an* R-approximator *if the following holds:*

- *Z is compatible with translation and scaling.*
- *For each $k \geq 1$, for each normalized input (n, P), and for each spanning set Q^n for $\bar{B}^n$, there exists a multi-set[2] $Q \subseteq Q^n$ of size at most[3] k such that $\|Z(Q) - z^n\| = \|Z(Q)\| \leq R(k, n)$.*

The following meta-result is fairly obvious from the above discussions:

Lemma 5.11 *If Z is an R'-approximator and $R'(k, n) \leq R(k, n)$ for each pair (k, n), then $(A_k^{Z,R})$ is R-successful.*

[2] Replace "multi-set" by "set" if the generic algorithm is run in NR-mode.

[3] Replace "at most" by "exactly" if the generic algorithm is run in FSM-mode.

Let us briefly summarize what we have achieved so far. Setting $R' = R$, Lemma 5.11 implies that a Z-function which is an R-approximator leads to an instantiation $(A_k^{Z,R})$ of the generic algorithm that is R-successful. Lemma 5.1 states that an R-successful family of algorithms (for an admissible function R) can be used to solve the μ-relaxation of DCB in polynomial time (for each $\mu > 0$). In order to prove the main result of this section, Theorem 5.2, it is sufficient to analyze the functions $Z_{\mathrm{CG}}, Z_{\mathrm{SB}}, Z_{\mathrm{ED}}$ and to prove that each of them is an R-approximator for some admissible function R. This is exactly what we will do in the course of the next two subsections.

5.3 The Center-of-Gravity Algorithm

The analysis of the Center-of-Gravity algorithm builds on Theorem 5.12, which is attributed to Maurey [7]. A proof for this theorem can be found in [11, 5] or in [1].

Theorem 5.12 (Maurey, [1]) *Let F be a vector space with a scalar product $(\cdot,\cdot)$ and let $\|f\| \triangleq \sqrt{(f,f)}$ be the induced norm on F. Suppose $G \subseteq F$ and that, for some $c > 0$, $\|g\| \le c$ for all $g \in G$. Then for all f from the convex hull of G and all $k \ge 1$ the following holds:*

$$\inf_{g_1,\ldots,g_k \in G} \left\| \frac{1}{k} \sum_{i=1}^{k} g_i - f \right\| \le \sqrt{\frac{c^2 - \|f\|^2}{k}}.$$

The proof makes use of the probabilistic method. It is essential for the validity of the theorem that the elements $g_1, \ldots, g_k$ taken from G in the inf-expression are not necessarily different.

Corollary 5.13 *Z_{CG} is a $\sqrt{1/k}$-approximator.*

Proof: We apply Theorem 5.12 to the following special situation:

- $F = \Re^n$ with the standard scalar product. The induced norm is the Euclidean norm.
- $G = Q^n$, where Q^n is a spanning set for the unit ball $\bar{B}^n$. Remember that $|Q^n| \le n+1$ and all points of Q^n reside on the unit sphere S^n. Moreover, the convex hull of Q^n is the simplex induced by Q^n, which contains the center z^n of $\bar{B}^n$ (the origin).
- Choose $f = z^n$ (the origin). It follows that $\|f\| = 0$. Since $G = Q^n \subset S^n$, the bound c in Theorem 5.12 can be safely set to 1 in our particular application. Thus the upper bound given in the theorem simplifies to $\sqrt{1/k}$.

It follows from this discussion that there exist points $g_1, \ldots, g_k$ taken from Q^n (possibly with repetitions) such that their center of gravity has distance at most $\sqrt{1/k}$ from the origin. Thus, Z_{CG} is a $\sqrt{1/k}$-approximator. □

5.4 The Smallest-Ball and the Equal-Distance Algorithm

Throughout this subsection, Q^n denotes a spanning set for the closed unit ball $\bar{B}^n$. Remember that Q^n consists of (at most $n+1$) points in general position that reside on the unit sphere S^n. The simplex $\mathcal{S}(Q^n)$ induced by Q^n contains the origin z^n. By abuse of notation, we will identify Q^n with $\mathcal{S}(Q^n)$.

Each subset Q of Q^n of size k induces the sub-simplex of Q^n with vertex set Q, briefly called a *k-sub-simplex* hereafter. Again, we identify Q with the induced k-sub-simplex. An n-sub-simplex is called a *face*, a 2-sub-simplex is called an *edge*, and a 1-sub-simplex is called a *vertex* of the simplex Q^n.

Let Q be a face of Q^n and H_Q its supporting hyper-plane. Note that the intersection of H_Q and $\bar{B}^n$ is a closed $(n-1)$-dimensional ball, say $\bar{B}_Q$. Its center and its boundary are denoted as z_Q and S_Q, respectively. Note that Q is contained in S_Q, but is not necessarily a spanning set for $\bar{B}_Q$. We say that Q is a *central face* of Q^n if Q minimizes the distance between z_Q and z^n.[4] For such faces, the following holds:

Lemma 5.14 *Each central face Q of Q^n is a spanning set for $\bar{B}_Q$.*

Proof: Figure 6 shows the simplex Q^n with a face Q which is not a spanning set for $\bar{B}_Q$. Clearly, z_Q is the projection of z^n onto H_Q. Since z_Q does not belong to the face Q, the line connecting z^n and z_Q penetrates another face. It follows that the center of the penetrated face is closer to the origin z^n than z_Q. Thus, Q is not a central face. ∎

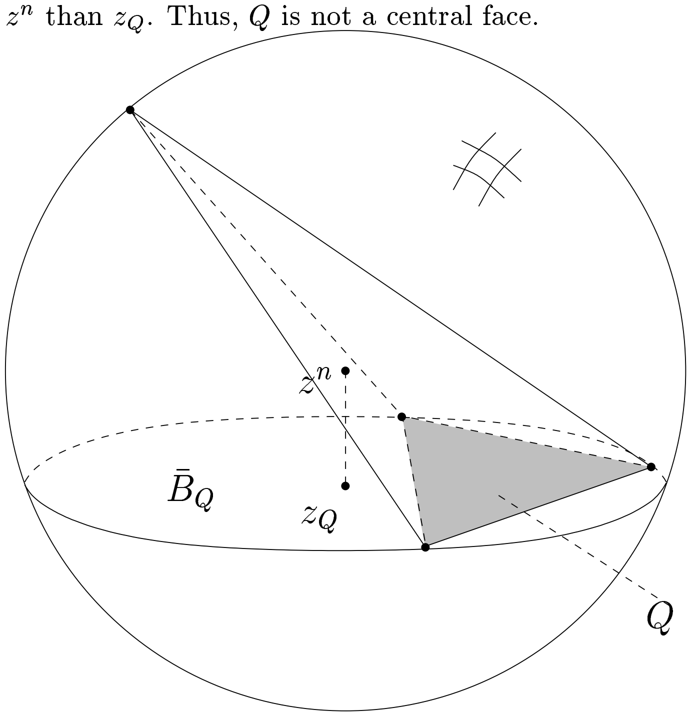


Figure 6: A non-degenerated spanning set Q^3 for $\bar{B}^3$ with a face Q that is not a spanning set for the corresponding sub-ball $\bar{B}_Q$.

These considerations can easily be generalized from faces to arbitrary sub-simplexes. We first generalize the notion of central faces by means of downward induction:

- The notion of a central n-sub-simplex of Q^n has already been defined (because a central n-sub-simplex is a central face).
- A k-sub-simplex of Q^n is called *central* if it is the central face of a central $(k+1)$-sub-simplex of Q^n.

[4] Since z^n is the all-zeroes vector, this distance coincides with $\|z_Q\|$.

Let Q be a k-sub-simplex of Q^n, and let U_Q be the $(k-1)$-dimensional affine sub-space of $\Re^n$ that contains Q. The intersection of U_Q and $\bar{B}^n$ is a closed $(k-1)$-dimensional ball, say $\bar{B}_Q$. Its center and its boundary are denoted as z_Q and S_Q, respectively. Q is contained in S_Q, but is not necessarily a spanning set for $\bar{B}_Q$. The following result easily follows by induction:

Lemma 5.15 *Each central k-sub-simplex Q of Q^n is a spanning set for $\bar{B}_Q$.*

Note furthermore that there exists at least one central k-sub-simplex Q of Q^n for all $k = 0, \ldots, n$.

Assume that Q is a k-sub-simplex of Q^n which is a spanning set for $\bar{B}_Q$. From Corollary 5.8, we know that $Z_{\mathrm{SB}}(Q) = Z_{\mathrm{ED}}(Q) = z_Q$. According to Lemma 5.15, this equality holds in particular for each central k-sub-simplex. The analysis of the Smallest-Ball and the Equal-Distance algorithm is based on the following

Lemma 5.16 *Let Q be a central k-sub-simplex of Q^n. Then $\|z_Q\| \le R_2(k,n)$, where $R_2(k,n)$ is the function given by equation (7).*

Lemma 5.16, whose (somewhat lengthy) proof is given in the appendix, basically concludes the analysis of the Smallest-Ball and the Equal-Distance algorithm:

Corollary 5.17 *Z_{SB} and Z_{ED} are both R_2-approximators.*

Proof: Let Q^n be a spanning set for $\bar{B}^n$. Let Q be a central k-sub-simplex of Q^n. According to Lemma 5.15, Q is a spanning set for $\bar{B}_Q$. It follows that $Z_{\mathrm{SB}}(Q) = Z_{\mathrm{ED}}(Q) = z_Q$. According to Lemma 5.16, $\|z_Q\| \le R_2(k,n)$. □

6 Conclusions

We briefly mention some possible extensions of our work and some open questions:

- All hardness results presented in this paper remain true when we disallow multi-sets and consider only points of "multiplicity" 1. The proofs would become technically more involved[5] (without providing much more insight).
- The notion of μ-relaxation can be generalized (in the obvious fashion) from a constant μ to a function μ in parameters n (the dimension) or m (the number of points in the input instance).
- It can be shown [4] that the $\sqrt{1/(45n)}$-relaxation of DOH and the $1/(90n)$-relaxation of DOB (or DCB) are NP-hard (and, in fact, even NP-hard to approximate). On the other hand, we have shown in this paper that the μ-relaxation of these problems can be solved in polynomial time for each constant $\mu > 0$. These results leave open the computational complexity of the μ-relaxation of DOH (or DOB,DCB, respectively), when $\mu = \mu(n)$ approaches zero asymptotically slower than $\sqrt{1/n}$ (or $1/n$, respectively).
- The NP-hardness of the $1/(90n)$-relaxation of DCB shows that we cannot expect an R-successful algorithm for DCB — for an admissible function $R = R(k,n)$ — with a polynomial time bound in k. However, the time bound $O(m^k)\mathrm{poly}(n,m)$ achieved by our

[5]Basically, an input point with multiplicity j must be replaced by j different points with approximately the same location in $\Re^n$.

algorithms $A_{\mathrm{CG}}, A_{\mathrm{SB}}, A_{\mathrm{ED}}$ might be improved to time bounds of the form $f(k)\mathrm{poly}(n, m)$ for some function f. In the parameterized complexity framework [6], this is the question of whether the $(1/k)$-relaxation of DCB is fixed-parameter tractable.

- In this paper, we investigated the problem of maximizing the empirical density (as opposed to the true density w.r.t. an input generating distribution). From this purely combinatorial perspective, the Center-of-Gravity algorithm is almost as successful as the (computationally more expansive) Equal-Distance algorithm (not to speak of the even more expansive Smallest-Ball algorithm). We expect however A_{ED} and A_{SB} to exhibit a superior statistical generalization. The validation of this claim will be the subject of future research.

Acknowledgment

The authors gratefully acknowledge the support of the German-Israeli Foundation for Scientific Research and Development Grant I-403-001.06/95. This work has been supported in part also by the ESPRIT Working Group in Neural and Computational Learning II, NeuroCOLT2, No. 27150. Thanks to Robert C. Williamson for pointing our attention to the Theorem of Maurey. Thanks to Dietrich Braess and Rüdiger Verfürth for helpful discussions concerning the proof of Lemma 5.16.

References

[1] Martin Anthony and Peter L. Bartlett. *Neural Network Learning: Theoretical Foundations.* Cambridge University Press, 1999.

[2] Shai Ben-David, Nadav Eiron, and Philip Long. On the difficulty of approximately maximizing agreements. In *Proceedings of the 13th Annual Conference on Computational of Learning Theory*, pages 266–274, 2000.

[3] Shai Ben-David and Michael Lindenbaum. Learning distributions by their density levels – a paradigm for learning without a teacher. *Journal of Computer and System Sciences*, 55(1):171–182, 1997.

[4] Shai Ben-David and Hans Ulrich Simon. Efficient learning of linear perceptrons. In *Advances in Neural Information Processing Systems 13*, 2001. To appear.

[5] B. Carl. On a characterization of operators from l_q into a Banach space of type p with some applications to eigenvalue problems. *Journal of Functional Analysis*, 48:394–407, 1982.

[6] Rodney G. Downey and Michael R. Fellows. *Parameterized Complexity.* Springer Verlag, 1998.

[7] R. M. Dudley. Universal Donsker classes and metric entropy. *Annals of Probability*, 15(4):1306–1326, 1987.

[8] Johan Håstad. Some optimal inapproximability results. In *Proceedings of the 29th Annual Symposium on Theory of Computing*, pages 1–10, 1997.

[9] David S. Johnson and F. P. Preparata. The densest hemisphere problem. *Theoretical Computer Science*, 6:93–107, 1978.

[10] Nimrod Megiddo. On the complexity of polyhedral separability. *Discrete Computational Geometry*, 3:325–337, 1988.

[11] G. Pisier. Remarques sur un résultat non publié de B. Maurey. *Séminaire d'Analyse Fonctionelle 1980–1981*, Ecole Polytechnique, Centre de Mathématiques, Palaiseau, V.1–V.12, 1981.

A Proof of Lemma 5.16

Remember that Q^n denotes a spanning set for the unit ball $\bar{B}^n$. Let Q be a central k-sub-simplex of Q^n. Remember that the intersection of $\bar{B}^n$ with the lowest dimensional affine sub-space of $\Re^n$ that contains Q yields a $(k-1)$-dimensional ball B_Q with center z_Q. We have to show that $\|z_Q\| \le R_2(k,n)$, where R_2 is the function given by equation (7).

If Q^n is a degenerated spanning set for $\bar{B}^n$, then it is a non-degenerated spanning set for a lower-dimensional unit ball, say for $\bar{B}^{n'}$ such that $n' < n$. Since $R_2(k,n') < R_2(k,n)$ if $n' < n$, we may restrict ourselves to the non-degenerated case in what follows. Thus, Q^n consists of $n+1$ vertices, say $q_0, \dots, q_n$, that are in general position and reside on the unit sphere S^n. Viewed as simplex, Q^n contains the origin.

We apply induction on $n+1-k$. The case $k = n+1$ (induction base) is trivial. The only $(n+1)$-sub-simplex of Q^n is Q^n itself. Thus z_Q coincides with the origin z^n. Therefore, $\|z_Q\| = \|z^n\| = 0 = R_2(k, n+1)$. The following result covers the case $k = n$.

Lemma A.1 *Let Q be a central face of Q^n. Then, $\|z_Q\| \le 1/n$.*

Proof: We will derive several formulas for the (n-dimensional) volume $\mathcal{V}$ of the simplex Q^n (in terms of $\|z_Q\|$ and some other parameters) which algebraically imply that $\|z_Q\| \le 1/n$.

Let $Q_i = Q^n \setminus \{q_i\}$. Remember that, by abuse of notation, Q_i denotes also the face induced by the vertices of $Q^n \setminus \{q_i\}$. Let $\mathcal{V}_i$ denote the ($(n-1)$-dimensional) volume of Q_i. Let Q'_i be the simplex that is obtained from Q^n when we replace vertex q_i by the origin, and let $\mathcal{V}'_i$ denote the (n-dimensional) volume of Q'_i. Let finally h_i denote the distance between vertex q_i and face Q_i (i.e., the height of Q^n when viewed as simplex on top of face Q_i), and let r_i denote the distance between the origin and Q_i (i.e., the height of Q'_i when viewed as simplex on top of face Q_i). An illustration of these notations may be found in Figure 7. Note that

$$\|z_Q\| = \min_{i=0,\dots,n} r_i \tag{8}$$

because Q is a central face.

We proceed with the following auxiliary result:

Claim A.2 *For all $i = 0, \dots, n$: $h_i \le 1 + r_i$.*

Proof of the Claim: Let z_i be the projection of the origin to face Q_i. Clearly, h_i is not greater than the distance from q_i to z_i, i.e., $h_i \le \|z_i - q_i\|$. As a vertex of Q^n, q_i has distance 1 from the origin, and (by definition of r_i) the origin has distance r_i to z_i. Using the triangle inequality, we conclude that $h_i \le \|z_i - q_i\| \le 1 + r_i$. □

We are now prepared to derive various formulas for $\mathcal{V}$. Remember that the n-dimensional volume of a simplex in $\Re^n$, viewed as simplex of height h_* on top of a face Q_* with $(n-1)$-dimensional volume $\mathcal{V}_*$, is given by $h_*\mathcal{V}_*/n$. In combination with Claim A.2, we get

$$\mathcal{V} = \frac{h_i \mathcal{V}_i}{n} \le \frac{(1+r_i)\mathcal{V}_i}{n}. \tag{9}$$

Summing over all i, we obtain

$$(n+1)\mathcal{V} = \tfrac{1}{n}(h_0\mathcal{V}_0 + \cdots + h_n\mathcal{V}_n) \le \tfrac{1}{n}\left((1+r_0)\mathcal{V}_0 + \cdots + (1+r_n)\mathcal{V}_n\right). \tag{10}$$

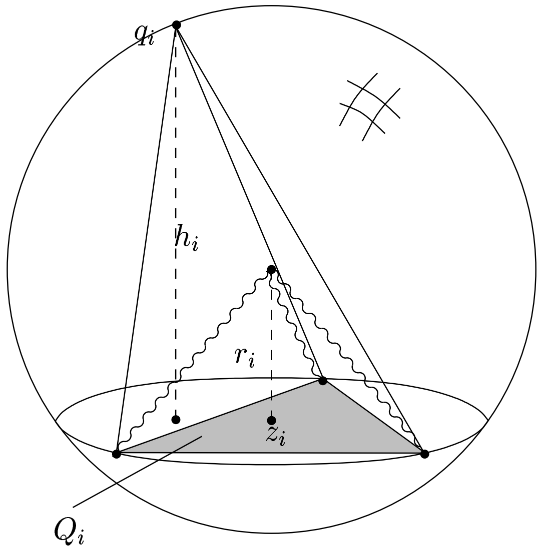


Figure 7: A non-degenerated spanning set Q^3 and the decomposition of the corresponding simplex into sub-simplexes. The serpentine lines indicate segments of length 1.

Since Q^n partitions into $Q'_0, \ldots, Q'_n$ (up to an overlap of n-dimensional volume zero), we may alternatively write $\mathcal{V}$ as follows:

$$\mathcal{V} = \mathcal{V}'_0 + \cdots + \mathcal{V}'_n = \frac{1}{n}(r_0\mathcal{V}_0 + \cdots + r_n\mathcal{V}_n) \tag{11}$$

Subtracting (11) from (10), we get

$$n\mathcal{V} \leq \frac{1}{n}(\mathcal{V}_0 + \cdots + \mathcal{V}_n) = \frac{\mathcal{O}}{n}, \tag{12}$$

where $\mathcal{O} = \mathcal{V}_0 + \cdots + \mathcal{V}_n$ is the ($(n-1)$-dimensional) volume of the surface of Q^n. Dividing (11) by (12), we obtain

$$\frac{1}{n} \geq \frac{\mathcal{V}_0}{\mathcal{O}} \cdot r_0 + \cdots + \frac{\mathcal{V}_n}{\mathcal{O}} \cdot r_n. \tag{13}$$

Note that the right hand of this inequality is a convex combination of $r_0, \ldots, r_n$ and therefore lower-bounded by $\min_{i=0,\ldots,n} r_i = \|z_Q\|$. This completes the proof of Lemma A.1. $\square$

As a marginal note, we would like to mention that (12) implies that $\mathcal{V}/\mathcal{O} \leq 1/n^2$. Quantity $1/n^2$ is precisely the volume-surface ratio of the regular simplex[6] with $n+1$ vertices residing on the unit sphere S^n. (Figure 8 shows the regular simplexes in $\Re^2$ and $\Re^3$, respectively.) We have therefore accidentally proven that regular simplexes (with vertices residing on the unit sphere) achieve the highest volume-surface ratio (among all simplexes whose vertices reside on the unit sphere).

We are now prepared to perform the inductive step. Let $Q = Q^{k-1}$ be a central k-sub-simplex of Q^n for some $k < n$. It follows that there is a chain

$$Q^{k-1} \subset \cdots \subset Q^{n-1} \subset Q^n,$$

[6] A simplex is called *regular* if all of its edges have the same length. By symmetry, the vertices of each regular simplex reside on the boundary of a ball around their center of gravity. The regular simplex is unique up to translation, rotation and scaling.

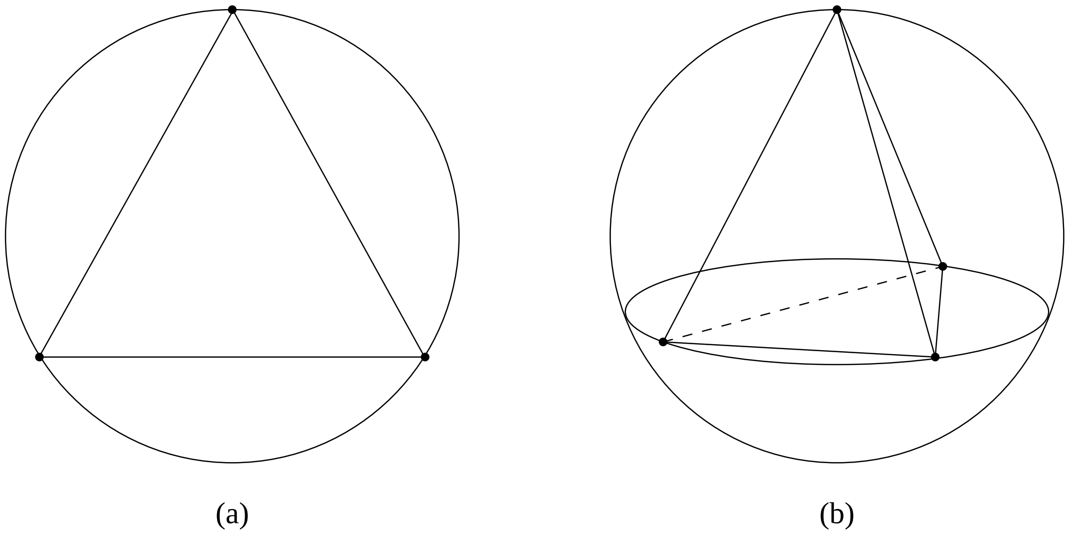


Figure 8: (a) A regular simplex with 3 vertices. (b) A regular simplex with 4 vertices.

such that Q^j is a $(j+1)$-sub-simplex of Q^n for $j = k-1, \ldots, n$, and Q^j is a central face of Q^{j+1} for $j = k-1, \ldots, n-1$. We denote the sub-ball of $\bar{B}^n$, obtained by intersecting $\bar{B}^n$ with the lowest-dimensional affine sub-space containing Q^j, as $\bar{B}(z_j, r_j)$. Here, z_j is the center of this sub-ball and r_j is the radius. Note that z_n coincides with z^n (the origin) and $r_n = 1$. Since $z_{n-1} = z_{Q^{n-1}}$ and Q^{n-1} is a central face of Q^n, Lemma A.1 yields

$$\|z_n - z_{n-1}\| = \|z_{n-1}\| \leq 1/n. \tag{14}$$

Furthermore, $z_{k-1} = z_Q$. We define $R(k,j) \stackrel{\triangle}{=} \|z_j - z_{k-1}\|$ for $j = k-1, \ldots, n$. Note that $R(k, k-1) = 0$ and $R(k,n) = \|z_{k-1}\| = \|z_Q\|$. It suffices therefore to show that $R(k,n) \leq R_2(k,n)$.

We may apply the induction hypothesis to Q as k-sub-simplex of Q^{n-1}, keeping in mind that $\bar{B}(z_{n-1}, r_{n-1})$ has radius r_{n-1} (and not radius 1). Taking the scaling factor r_{n-1} into consideration, the inductive hypothesis reads as follows:

$$R(k, n-1) \leq r_{n-1} \cdot R_2(k, n-1) \tag{15}$$

Let z_0 be a vertex in Q. The rest of the proof, which is illustrated in Figure 9, makes use of the fact that the triangles induced by z_0, z_{n-1}, z_n and z_{k-1}, z_{n-1}, z_n, respectively, have both a right angle at z_{n-1}. The Pythagorean Law, applied to both triangles, yields:

$$\begin{aligned} r_{n-1}^2 &= 1 - \|z_{n-1}\|^2 && (16)\\ R^2(k,n) &= R^2(k,n-1) + \|z_{n-1}\|^2 && (17) \end{aligned}$$

In combination with equations (15) and (14), we get

$$\begin{aligned} R^2(k,n) &\leq r_{n-1}^2 R_2^2(k,n-1) + \|z_{n-1}\|^2 \\ &= (1 - \|z_{n-1}\|^2) R_2^2(k,n-1) + \|z_{n-1}\|^2 \\ &\leq \left(1 - \frac{1}{n^2}\right) R_2^2(k,n-1) + \frac{1}{n^2} \\ &= R_2^2(k,n). \end{aligned}$$

The last inequality holds because $(1-\|z_{n-1}\|^2)R_2^2(k,n-1)+\|z_{n-1}\|^2$ is a convex combination of $R_2^2(k,n-1)$ and 1 and $R_2^2(k,n-1) \leq 1$. This convex combination is maximized when $\|z_{n-1}\|^2$ equals its upper bound $1/n^2$. The last equality (expressing $R_2(k,n)$ in terms of $R_2(k,n-1)$) is obtained by a straightforward calculation that uses equation (7). It follows from our calculations that $R(k,n) \leq R_2(k,n)$, which concludes the proof.

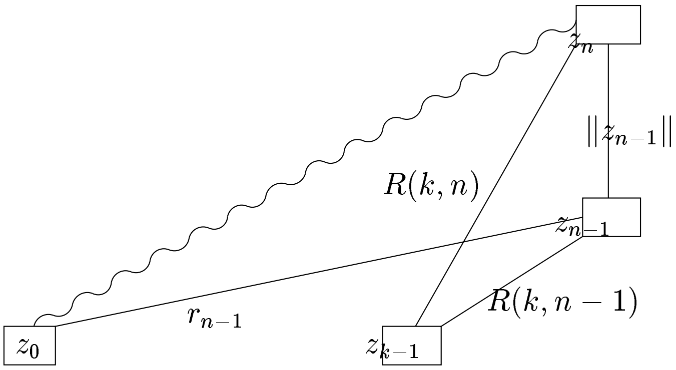


Figure 9: Two triangles (with a right angle at z_{n-1}, respectively) induced by the center z_n of an n-dimensional simplex Q^n and the centers of some sub-simplexes of Q^n. Serpentine lines indicate segments of length 1.

Finally, we would like to mention that the upper bound on $\|z_Q\|$ given in Lemma 5.16 is tight: if Q is a k-sub-simplex of the regular simplex with $n+1$ vertices residing on S^n, then $\|z_Q\| = R_2(k,n)$. We omit the straightforward inductive proof of this claim.