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

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 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 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 overfitting. 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 $\mathbf{R}^\mathbf{n}$. 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 [2]. 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, [2] 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 S of the domain – output a set $h \in \mathcal{H}$ that maximizes $|S \cap h|$*

We consider two hypothesis classes: the class of axis aligned hypercubes and the class of balls (both in $\mathbf{R}^\mathbf{n}$). For each of these classes we prove that there exists some $\gamma > 0$ (independent of the input sample size and dimensionality) such that, unless P=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 an alternative to the common definition of approximation. 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 $\mu > 0$,

> a μ-*successful* learning algorithm for $\mathcal{H}$ is an algorithm that, for every input sample, outputs a member of $\mathcal{H}$ that contains as many sample points as any member of $\mathcal{H}$ can contain *with margin* $> \mu$ (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, A μ-successful 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-success 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 [3] we show that the class of linear perceptrons is another such case.

This paper investigates the existence of μ-successful learning algorithms. Clearly, μ-success gets easier to achieve as μ grows, and is hardest for $\mu = 0$, in which case it becomes the usual optimization problem (without approximation). These exact optimization problems – finding the densest ball or the densest hypercube – are NP-hard, as we show below (for other NP-hardness results of this type see [5], [6]). We are interested in determining the values of μ at which the NP hardness of the approximation 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 μ-successful algorithms for the class, while on the other hand, for every $\mu < \mu_0$ μ-learnability is NP-hard. That is, assuming P$\neq$NP, no μ-successful learning algorithms for the class runs in polynomial time. Furthermore, there exists a positive constant such that for every $\mu < \mu_0$ it is NP-hard to approximate the optimal margin-μ success ratio of members of the class to within this constant factor. A similar phenomena holds also for the class of linear perceptrons [3].

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. Next we discuss the class of balls, Section 4 brings the hardness result for this class while the following Section 5 provides the optimization algorithm for the μ-margin 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 this 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 the previously known hardness-of-approximation results that we shall employ in our work.

2.1 THE COMBINATORIAL OPTIMIZATION PROBLEMS

In this paper, 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, S) where S 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 S as possible (accounting for their multiplicity in S).

We shall mainly be concerned with two instantiations of the above problem;

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

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 $I_i = [a_i, a_i + 1]$.

In both cases let us denote $m = |S|$.

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

MAX-E2-SAT 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. We denote by m the number of clauses in C.

BSH Inputs are of the form (n, S_+, S_-), where $n \geq 1$, and S_+, S_- are multi-sets of (not necessarily different) points from $\Re^n$. We denote $m_+ = |S_+|$ and $m_- = |S_-|$. A hyper-plane $H(w,t)$, where $w \in \Re^n$ and $t \in \Re$, *correctly classifies* $p \in S_+$ if $wp > t$, and it *correctly classifies* $p \in S_-$ if $wp < t$. The problem is to find the *Best Separating Hyper-plane* for S_+ and S_-, that is, a pair $(w,t) \in \Re^n \times \Re$ such that $H(w,t)$ correctly classifies as many points from $S_+ \cup S_-$ as possible (accounting for their multiplicities in the lists S_+ and S_-).

DOH This is the densest set approximation problem for the class of open hemispheres. That is, inputs are multi-sets S of points from S^{n-1} - the $(n-1)$ dimensional unit sphere. and each class $\mathcal{H}_n$ is the class of all sets of the form $\{x : wx > 0\}$ for $w \in \Re^n$.

DCB The *Densest Closed Ball* problem is the same as the Densest Open Ball problem, except that now $\mathcal{H}_n$ consists of closed radius 1 balls.

2.2 BASICS OF COMBINATORIAL OPTIMIZATION APPROXIMATION THEORY

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

$$\frac{\text{opt}(I) - A(I)}{\text{opt}(I)}$$

is called the *relative error of A on input instance I*. A is called a *δ-approximation algorithm for* Π, where $\delta \in \Re^+$, if its relative error on I is at most δ for all input instances I.

The tool we use to show hardness results for these maximization problems is the basic cost-preserving reduction: Let Π and Π' be two maximization problems. A *cost preserving polynomial reduction from* Π *to* Π', written as $\Pi \le^{\text{cp}}_{\text{pol}} \Pi'$ consists of the following components:

- a polynomial time computable mapping $I \mapsto I'$, which maps an input instance I of Π to an input instance I' of Π'
- for each I, a mapping $\sigma \mapsto \sigma'$, which maps a legal solution σ for I with profit s to a legal solution σ' for I' with the same profit s
- for each I, a polynomial time computable mapping $\sigma' \mapsto \sigma$, which maps a legal solution σ' for I' with profit s to a legal solution σ for I with the same profit s

The following result is evident:

Lemma 2.1 *If $\Pi \le^{cp}_{pol} \Pi'$ and there is no polynomial time δ-approximation algorithm for Π, then there is no polynomial time δ-approximation algorithm for Π'.*

2.3 SOME KNOWN HARDNESS-OF-APPROXIMATION RESULTS

We shall base our hardness reductions on two known results.

Theorem 2.2 [Håstad, [4]] *Assuming P≠NP, for any $\delta < 1/22$, there is no polynomial time δ-approximation algorithm for MAX-E2-SAT.*

Theorem 2.3 [Ben-David, Eiron and Long, [1]] *Assuming P≠NP, for any $\delta < 3/418$, there is no polynomial time δ-approximation algorithm for BSH.*

Claim 2.4 *BSH$\le^{cp}_{pol}$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 2.5 *Assuming P≠NP, for any $\delta < 3/418$, there is no polynomial time δ-approximation algorithm for DOH.*

Before we proceed, let us fix some notation for the subsets of $\Re^n$ that we shall be dealing with. Let $n \ge 1$, $w, z \in \Re^n$, $t \in \Re$, 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 halfspace, respectively. $B(z,R) = \{x \in \Re^n : ||z - x||_2 < R\}$ denotes the open ball of radius R around center z. $\bar{B}(z,R) = \{x \in \Re^n : ||z - x||_2 \le R\}$ denotes the corresponding closed ball. $S^{n-1} = \{x \in \Re^n : ||x||_2 = 1\}$ denotes the $(n-1)$-dimensional unit sphere.

2.4 THE NEW NOTION OF APPROXIMATION – μ-RELAXED DENSEST SET PROBLEMS

As mentioned in the introduction, we shall also discuss a variant of the above notion of approximation for densest set problems. The idea above this new notion, that we term 'μ-margin approximation', is that the required approximation rate varies with the structure of the inout sample. When there exist optimal solutions that are 'stable', in the sense that minor variations to these solutions will not effect 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.

Definition 2.6 *Given a hypothesis class $\mathcal{H} = \cup_n \mathcal{H}_n$ and a real parameter $\mu > 0$,*

- *For $h \subset \Re^n$, let $h^{-\mu}$ be the set of points that are included in h with a margin μ. That is:*

$$h^{-\mu} \triangleq \{x \in \Re^n \ : \ \bar{B}(x,\mu) \subseteq h\}$$

- *Given a finite $P \subset \Re^n$, a hypothesis $h \in \mathcal{H}_n$ is a μ-*margin approximation *for P w.r.t. $\mathcal{H}$ if*

$$|P \cap h| \ge \max_{h \in \mathcal{H}} |P \cap h^{-\mu}|.$$

- *The μ-*relaxation *version of the densest set problem for a class $\mathcal{H}$ is to output, for every n and every finite input set $P \subset \Re^n$, a μ-margin approximation for P w.r.t. $\mathcal{H}$. We shall use superscript μ to denote the μ-relaxed version of a problem. For example, DAC$^\mu$ denotes the μ relaxation of DAC.*
- *A densest set algorithm A for a class $\mathcal{H}$ is an algorithm that on input P $subset \Re^n$ outputs a hypothesis $A(P) \in \mathcal{H}_n$.*
- *A densest set algorithm A is μ-*successful *for $\mathcal{H}$ if it solves the μ-relaxation version of the densest set problem for $\mathcal{H}$. In other words, if for every finite input $P \subset \Re^n$, its output $A(P)$ is a μ-margin approximation for P w.r.t. $\mathcal{H}$.*

3 THE DENSEST CUBE PROBLEM AND ITS μ-RELAXATIONS

For the DAC problem, we show a hardness of approximation result, that extends to the μ-relaxed problem for every $\mu < 1/4$. We complement this result with a positive result, showing that there exist a μ-succesful polynomial time algorithm for DAC, for all $\mu \geq 1/4$. This result is quite surprising in that it gives a tight bound on the relaxation required to solve DAC efficiently. Let us first show the negative result:

Theorem 3.1 *The densest cube problem is NP-hard to approximate to within ϵ for every $\epsilon < 1/22$. Furthermore, this hardness result holds for the relaxed DAC$^\mu$ problem for any $\mu < 1/4$.*

Proof: We define a cost-preserving reduction of MAX-E2-SAT to the densest cube problem. First, we define a mapping ϕ from instances of MAX-E2-SAT (over n variables) to subsets of $\Re^n$. Let $v_1 \ldots, v_n$ be the variables that appear in the propositional formulas. Let h be the following function:

$$\begin{aligned}\forall i \quad h(v_i) &= e_i &= (\underbrace{0,\ldots,0}_{i-1},1,0,\ldots,0)\\ \forall i \quad h(\overline{v_i}) &= -e_i &= (\underbrace{0,\ldots,0}_{i-1},-1,0,\ldots,0)\end{aligned}$$

Given a 2-clause $l_1 \vee l_2$, define

$$\phi(l_1 \vee l_2) \triangleq \{h(l_1)+h(l_2), -h(l_1)+h(l_2), h(l_1)-h(l_2)\}\,.$$

Informally, we associate with each clause three points in the hyper-plane spanned by the coordinates that correspond to the variables of this clause. These three points correspond to the three truth assignment to these variables that satisfy the given clause.

Finally, define ϕ over formulas to be:

$$\phi\left(\bigwedge_{i=1}^{m} c_i\right) = \bigcup_{i=1}^{m} \phi(c_i).$$

Given a cube $R = \Pi_{i=1}^n I_i$ and a 2CNF formula ϕ, we define a truth assignment $g(R)$ by setting

$$g(C)(v_i) = T \textit{ iff } 1 \in I_i$$

In the other direction, given a truth assignment

$$s : \{v_1,\ldots,v_n\} \mapsto \{F,T\},$$

we define a corresponding cube $R(s) = \Pi t_i^s$ where

$$t_i^s = [0,1] \text{ if } s(v_i)=T \text{ and } t_i^s = [-1,0] \text{ if } s(v_i)=F$$

For any given $\mu < 1/4$ we constract a mapping ϕ_μ from propositional formulas to subsets of $\Re^n$. The mapping ϕ_μ is the same as the mapping ϕ above, except that for some $1-2\mu > \rho > 1/2$ all the coordinates of the points in $\phi(\alpha)$ are multiplied by ρ (so the values of the non-zero entries are $+/-\rho$ rather than $+/-1$). Similarly, we define the functions from cubes to truth assignments by $g_\mu(C) = T$ iff $\rho \in I_i$, and the functions in the inverse direction by $t_i^s = [-\mu, 1-\mu]$ if $s(v_i) = T$ and $t_i^s = [-(1-\mu), \mu]$ if $s(v_i) = F$. The following three observations suffice to show that, for every $\mu < 1/4$ these reductions are cost preserving.

In the following let α denote a 2-CNF formula and let $\phi(\alpha)$ be its image under the above construction.

1. For every α and every clause C in it, any unit-size rectangle contains at most one point from $\phi_\mu(C)$.
2. For every unit-size rectangle h and every clause C, if $h \cap \phi_\mu(C) \neq \emptyset$ then $g(h)$ satisfies C.
3. For every clause C and every truth assignment s, if s satisfies C then $R(s) \cap \phi(C) \neq \emptyset$.

∎

The complementary positive result uses a rather simple algorithm. Recall that all that is required of a μ-successful algorithm for the DAC problem, is to output a cube C that includes as many points of the input as an optimal cube would include with a margin at least μ. Let $\{p_1,\ldots,p_m\}$ be the points in P, the input to the algorithm.

Algorithm 3.2:

1. For all $i \in \{1,\ldots,m\}$ do:
   (a) Let C_i be the cube of side length 1 whose center is p_i.
   (b) Let $c_i \triangleq |C_i \cap P|$.
2. Output the cube C_j for which c_j is maximal.

Theorem 3.3 *Algorithm 3.2 is a polynomial time μ-successful algorithm for DAC for any $\mu \geq 1/4$.*

Proof: The claim on the running time of the algorithm is immediate: We loop over all points, construct a cube around each, and check inclusion of all other points within this cube. All these operations may easily be carried out in time polynomial in n and m.

To prove the correctness of the algorithm, we have to show that for any $\mu \geq 1/4$, the output of this algorithm is a cube containing at least as many points from P as the optimal cube of side length 1 does. Let C be some cube of side length 1, that includes a maximal number of points from P with a margin μ (i.e., $|P \cap C^{-\mu}|$ is maximal). Consider some $p_i \in C \cap P$. The cube C_i, defined above in the description of the algorithm, will contain all points that have distance $\leq 1/2$ from p_i in the l_∞ norm. However, no pair of points in $P \cap C^{-\mu}$ are more than $1/2$ apart in the l_∞ norm. Hence, $P \cap C^{-\mu} \subseteq C_i \cap P$. Since the algorithm chooses the C_i that maximizes the number of points it includes, its output will satisfy the conditions required from a μ-successful algorithm. ∎

4 HARDNESS OF APPROXIMATION FOR THE DENSEST BALL PROBLEM

In this section, we call $H_+(w,0)$ an *open hemisphere* because we use the hyper-plane $H(w,0)$ as a separator of the unit sphere S^{n-1} into two hemispheres. We may furthermore assume that $||w||_2 = 1$ because vectors w and λw, $\lambda > 0$, lead to the same separation of S^{n-1}.

Lemma 4.1 *DOH$\leq_{pol}^{cp}$DOB.*

Proof: Let P be an input instance for DOH of size m in S^{n-1}. We choose the same P as the corresponding input instance I' of DOB.

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

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

$$\forall z \in \Re^n : \quad B(z,1) \subseteq H_+(z,0). \qquad (2)$$

These statements imply that there exists an open hemisphere containing at least s points of P iff there exists an open unit ball containing at least s points of P. Thus, we have a cost-preserving reduction from DOH to DOB.

To prove statement (1), we set $\mu = \min_{p \in C(w,P)} |wp|$. This implies that $wq \geq \mu$ 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 $ww = qq = 1$, this claim is evident from the following calculation:

$$\begin{aligned} ||z-q||_2^2 &= (z-q)(z-q) \\ &= zz - 2zq + qq \\ &= \mu^2 ww - 2\mu wq + qq \\ &= \mu^2 - 2\mu wq + 1 \\ &\leq \mu^2 - 2\mu^2 + 1 \\ &= 1 - \mu^2 \\ &< 1 \end{aligned}$$

In order to prove statement (2), we have to show that each $q \in B(z,1)$ satisfies $zq > 0$. To this end, note first that $q \in B(z,1)$ implies that

$$\begin{aligned} 1 > ||z-q||_2^2 &= zz - 2zq + qq \\ &= zz - 2zq + 1 \\ &\geq -2zq + 1. \end{aligned}$$

Clearly, this implies that $zq > 0$. ∎

Applying Corollary 2.5 we readily get

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

Applying a similar construction we get a hardness result for the densest set problem for the class of *closed* radius 1 balls. Namely,

Theorem 4.3 *Assuming P≠NP, for any $\delta < 1/198$, there is no polynomial time δ-approximation algorithm for DCB.*

For lack of space we defer the proof to the full version of the paper.

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 set of points in $\Re^n$. In this section, we show that, for each constant $\mu > 0$, the μ-relaxation of this problem can be solved optimally in polynomial time. Remember that, in the relaxed version of the problem, we assume that the input set $S \subset \Re^n$ is μ-robust for balls. It is easy to see that the μ-robustness of S is equivalent to the following condition:

If the optimal ball of radius 1 contains q points of S, then there exists a ball of radius $1-\mu$ which also contains q points of S.

The main result of this section is:

Theorem 5.1 *There exists a family $(A_k)_{k\geq 1}$ of polynomial time algorithms such that for all $n \geq 2$ the following holds. A_k on input $S \subseteq \Re^n$ outputs a point $y \in \Re^n$ such that the closed ball $\bar{B}(y, 1+\sqrt{1/k})$ contains not less points of S than the optimal closed ball of radius 1.*

We postpone the proof of this theorem to Subsections 5.1 and 5.2. Theorem 5.1 immediately implies the following result:

Corollary 5.2 *For each $\mu > 0$, there exists a polynomial time algorithm A which solves the μ-relaxation of the Densest Closed Ball Problem optimally.*

Proof: The proof is based on "Scaling and Rescaling". We omit the details in this brief abstract and only note that it is sufficient to choose k such that $\sqrt{1/k} \leq \frac{\mu}{1-\mu}$, to apply A_k to a properly scaled version of input S and to re-scale the output properly. $k = 1 + \lceil 1/\mu^2 \rceil$ is a possible choice for k. ∎

In the next two subsections, we describe and analyze the family A_k of algorithms that witnesses the correctness of Theorem 5.1.

5.1 A FAMILY OF ALGORITHMS FOR THE DENSEST BALL PROBLEM

The naive implementation of algorithm A_k is quite simple to describe. On input $S \subseteq \Re^n$, it exhaustively searches through all subsets T of S of size k and outputs the center z_T of the smallest ball in $\Re^n$ that contains T. Then it computes the profit $p(T)$ achieved by T, that is, the number of points from S that fall into the ball $\bar{B}(z_T, 1+\sqrt{1/k})$. Finally, it selects the set T with a maximal profit and outputs z_T. This implementation takes $|S|^k\text{poly}(|S|)$ steps under the unit cost measure. We briefly note that there exists a more clever implementation which takes only $2^{O(k)}\text{poly}(|S|)$ steps under the uniform cost measure. The interested reader is referred to the full paper.

In order to prepare the geometric analysis that is given in the next subsection, we make the following technical observations:

- We may assume without loss of generality that the optimal ball $\bar{B}_S$ of radius 1 for input S has at least $n+1$ points of S on its sphere. We briefly sketch why. First, we can conceptually shrink $\bar{B}_S$ as long as this does not diminish the profit. Let $S' \subseteq S$ denote the set of points from S belonging to $\bar{B}_S$ after the shrinking phase. Now $\bar{B}_S$ is the smallest ball containing S'. If $|S'| < n+1$, our analysis may switch to a lower-dimensional ball of dimension $|S'|-1$. Now $|S'| = n'+1$ holds for the reduced dimension n'. It can be strictly argued that the worst-case for the algorithm occurs, when no shrinking occurs and the dimensionality does not decrease. Details are given in the full paper.

- If $|S'| \geq n+1$ and $\bar{B}_S$ is the smallest ball containing S', then there exists a subset S'' of S' of size $n+1$ such that $\bar{B}_S$ is the smallest ball containing S''. The latter condition is equivalent to saying that the $(n+1)$-simplex induced by S'' contains the center of $\bar{B}_S$.

The main result of the next subsection, Corollary 5.11, will imply that there exists a subset T of S'' of size k such that $\bar{B}(z_T, 1+\sqrt{1/k})$ contains $\bar{B}_S$. This proves Theorem 5.1.

Since the location of the ball $\bar{B}_S$ is immaterial for our analysis, we may assume for the sake of simplicity that its center coincides with the origin $\bar{0} \in \Re^n$. Thus the analysis of algorithm A_k boils down to a purely geometric question: given a set S consisting of $n+1$ points on the sphere of the closed unit ball $\bar{B}(\bar{0}, 1)$ such that the induced $(n+1)$-simplex contains the origin, does there exist a subset T of S of size k such that $\bar{B}(z_T, 1+\sqrt{1/k})$ contains $\bar{B}(\bar{0}, 1)$? This question will be answered affirmatively in the next subsection.

5.2 GEOMETRIC ANALYSIS

Throughout this subsection, we make the following notational conventions. $\bar{0}$ denotes the all-zeros vector in $\Re^n$. $B_* = \bar{B}(\bar{0}, 1)$, i.e., B_* denotes the closed unit ball in $\Re^n$. S_* denotes the unit sphere in $\Re^n$. A *$(n+1)$-simplex with vertices on S^** is given by $n+1$ points (also called "vertices") from S_*. Since we discuss only $(n+1)$-simplexes with vertices on S_* (together with their sub-simplexes), we simply say $(n+1)$-simplex in the sequel. It should be clear without saying that each $(n+1)$-simplex contains a whole hierarchy of lower-dimensional sub-simplexes. A sub-simplex with k points is called *k-sub-simplex*. A n-sub-simplex is called a *face*, a 2-sub-simplex is called an *edge*, and a 1-sub-simplex is called a *vertex*. Let K be a $(n+1)$-simplex and K' a face of K. The unique hyper-plane in $\Re^n$ that contains K' is called the *supporting hyper-plane of* K' and denoted by $H(K')$ (or simply by H'). Note that the intersection of H' and B_* is a closed $(n-1)$-dimensional ball. We denote this ball by $B(K')$ (or simply B') and its center by $z(K')$ (or simply z'). Point $z(K')$ is called the *center of* K'. The notion of a center is generalized to lower-dimensional sub-simplexes in the obvious way. A simplex that does not contain its center is called *marginal*. Figure 1 shows two examples. In our proof we will also be interested in the radius of the ball inscribed in a simplex K. We denote the radius of such ball $r(K)$. $r(K)$ is, theorefore, the minimal distance between the origin and any point on a face if the $(n+1)$-simplex K.

A *regular $(n+1)$-simplex* is a $(n+1)$-simplex with the property that for all $k = 1, \ldots, n$, all its k-sub-simplexes are congruent. Since it is unique up to rotation, we speak of "the" regular $(n+1)$-simplex and denote it by K_*. All sub-simplexes of K_* are again regular (in a lower-dimensional space). Note that a center of a regular simplex coincides with the center of gravity of its vertices. Clearly, regularity implies non-marginality. Figure 2 shows the regular 3- and the regular 4-simplex (augmented with some additional information whose purpose will be clarified later). The serpentine lines indicate lines of length 1. The same convention is used for the subsequent figures.

Lemma 5.3 *Let z' be the center of a face K'_* of the regular $(n+1)$-simplex K_*. Then $\|z'\|_2 = 1/n$.*

Proof: The proof is illustrated in Figure 2(b). Note that z' is the projection of the origin $z = \bar{0}$ to the supporting hyperplane $H(K'_*)$. Choose an orthonormal base for $\Re^n$ with one base vector, say the i-th one, in direction (z, z'). Thus, z' has only one coordinate, namely z'_i, which is different from 0. Note that all of the n vertices in K'_* have z_i as i-th coordinate. The i-th coordinate of the remaining vertex is -1. Because of regularity, $z_i = 0$ is the average of the i-th coordinates of all $n+1$ vertices of K_*, i.e., $0 = z_i = -1 + nz'_i$. Thus, $z'_i = 1/n$. We conclude that $\|z'\|_2 = |z_i| = 1/n$. ∎

Lemma 5.4 *Let $z^{(k)}$ be the center of a k-sub-simplex K''_* of the regular $(n+1)$-simplex K_*. Then $\|z^{(k)}\|_2 = R_{k,n}$, where $R(k,n) = \sqrt{\frac{n-k+1}{kn}}$.*

Proof: The proof is illustrated in Figure 3(a).$z^{(n+1)} = \bar{0}$ denotes the center of K_*, $z^{(n)}$ denotes the center of the n-sub-simplex K'_* of K_* which contains K''_*, and $z^{(0)}$ denotes a vertex of K''_*. The rest of the proof makes use of the fact that the triangles induced by $z^{(0)}, z^{(n)}, z^{(n+1)}$ and $z^{(k)}, z^{(n)}, z^{(n+1)}$, respectively, have both a right angle at $z^{(n)}$.

Note first that K'_* is a regular n-simplex, except that its vertices belong to the sphere of a ball of radius $r_n = \|z^{(n)} - z^{(0)}\|_2 < 1$. According to the law of Pythagoras, $r_n^2 = 1 - 1/n^2$. Thus, K'_* is a regular n-simplex up to the scaling factor $r_n = \sqrt{1 - 1/n^2}$.

Let $R_{k,n} = \|z^{(k)}\|_2$, i.e., $R_{k,n}$ denotes the distance between $z^{(k)}$ and the center $z^{(n+1)} = \bar{0}$ of K_*. Note that $R'_{k,n-1} = \|z^{(n)} - z^{(k)}\|_2$ coincides with $R_{k,n-1}$ up to the scaling factor r_n, i.e., $R'_{k,n-1} = r_n R_{k,n-1}$. Applying again the law of Pythagoras, we arrive at the following recursion:

$$R_{k,n}^2 = \frac{1}{n^2} + (1 - 1/n^2) R_{k,n-1} \qquad (3)$$

Using the obvious fact that $R_{k,k-1} = 0$, an easy induction (to be presented in the full paper) shows that $R_{k,n} = \sqrt{\frac{n-k+1}{kn}}$ solves the recursion. ∎

Lemma 5.5 *The volume of the regular $(n+1)$-simplex K_* is larger than the volume of any non-regular $(n+1)$-simplex.*

Proof: The claim of the lemma is easy to establish for $n = 2$. Let now $n \geq 3$ and K be a non-regular $(n+1)$-simplex. An easy compactness argument shows that there exists a $(n+1)$-simplex with maximal volume. It is therefore sufficient to show that the volume of K is not maximal. Applying a transitivity argument, it is easy to see (although wrong for $n = 2$) that K must contain a non-regular face K'. The remainder of the proof is visualized in Figure 4. Let H' be the supporting hyper-plane of K', $z^{(0)}$ be the vertex of K outside K', and h be the distance between $z^{(0)}$ and H'. The volume of K can be written as vol$(K) = \frac{1}{2} \cdot h \cdot$ vol(K'). Compare with Figure 4(a).

If we replace K' by its *regularization*, i.e., the regular face within the $(n-1)$-dimensional ball $H' \cap B_*$, we obtain a new $(n+1)$-simplex whose volume exceeds the volume of K. This is because, according to our induction hypothesis, the regularization of K' has a higher volume than K', and the height-parameter h is left unchanged. Compare with Figure 4(b). $\Box$

Lemma 5.6 *Let K be a non-marginal $(n+1)$-simplex and K' be a face of K whose center z' has a minimal Euclidean norm (i.e., a minimal distance from the origin $z = \bar{0}$ which is the center of K). Then K' is non-marginal.*

Proof: Figure 5 shows a non-marginal $(n+1)$-simplex K with center $z = \bar{0}$ and a marginal face K' of K with center z'. Let H' be the supporting hyper-plane of K'. Clearly, z' is the projection of z onto H'. Since K' is non-marginal, z' does not belong to K'. Thus, the line connecting z and z' penetrates another face. It follows that the center of the penetrated face is closer to the origin z than z'. $\Box$

Definition 5.7 *Let K be a $(n+1)$-simplex whose center is $z^{(n+1)} = \bar{0}$. We will associate with K a sequence*

$$K_{n+1} \supset K_n \supset \cdots \supset K_0$$

with centers $z^{(n+1)}, z^{(n)} \ldots, z^{(0)}$, respectively, that is inductively defined as follows:

- $K_{n+1} = K$.
- *For all $k = n, \ldots, 0$, let K_d be a d-sub-simplex of K_{d+1} whose center $z^{(d)}$ has minimal distance from $z^{(d+1)}$ (ties broken arbitrarily).*

*A sequence that can be obtained in this way is called a K-*sequence.

It is obvious that a least one K-sequence exists. Furthermore, iterative application of Lemma 5.6 shows that all members of a K-sequence are non-marginal if K is non-marginal.

Lemma 5.8 *Each $(n+1)$-simplex K satisfies $r(K) \le 1/n$.*

In order to prove Lemma 5.8, we will derive several formulas for the (n-dimensional) volume V of K (in terms of $r(K)$ and some other parameters) which algebraically imply that $r(K) \le 1/n$.

Let $\mathcal{V} = \{v_0, \ldots, v_n\}$ be the set of vertices of K. Let K_i be the face of K that is spanned by $\mathcal{V} \setminus \{v_i\}$, and let V_i denote the ($(n-1)$-dimensional) volume of K_i. Let K_i' be the simplex that is obtained from K when we replace vertex v_i by the origin, and let V_i' denote the (n-dimensional) volume of K_i'. Let finally h_i denote the distance between vertex v_i and K_i (i.e., the height of K when viewed as simplex on top of face K_i), and let r_i denote the distance between the origin and K_i (i.e., the height of K_i' when viewed as simplex on top of face K_i). An illustration of these notations may be found in Figure 6. Note that the smallest r_i is the radius of the ball inscribed in K:

$$r(K) = \min_{i=0,\ldots,n} r_i \tag{4}$$

We proceed with the following auxiliary result:

Lemma 5.9 *For all $i = 0, \ldots, n$: $h_i \le 1 + r_i$.*

Proof: Let z_i be the projection of the origin to face K_i. Clearly, h_i is not greater than the distance from v_i to z_i, i.e., $h_i \le \|z_i - v_i\|_2$. As a vertex of K, v_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 $h_i \le \|z_i - v_i\|_2 \le 1 + r_i$. $\Box$

We are now prepared to derive various formulas for V. Remember that the n-dimensional volume of a simplex in $\Re^n$, viewed as simplex of height h_* on top of a face K_* with $(n-1)$-dimensional volume V_*, is given by $h_* V_*/n$. In combination with Lemma 5.9, we get

$$V = \frac{h_i V_i}{n} \le \frac{(1+r_i)V_i}{n}. \tag{5}$$

Summing over all i, we obtain

$$\begin{aligned}(n+1)V &= \tfrac{1}{n}(h_0 V_0 + \cdots + h_n V_n) \\ &\le \tfrac{1}{n}\left((1+r_0)V_0 + \cdots + (1+r_n)V_n\right).\end{aligned} \tag{6}$$

Since K partitions into $K_0', \ldots, K_n'$ (up to an overlap of n-dimensional volume zero), we may alternatively write V as follows:

$$V = V_0' + \cdots + V_n' = \frac{1}{n}(r_0 V_0 + \cdots + r_n V_n) \tag{7}$$

Subtracting (7) from (6), we get

$$nV \le \frac{1}{n}(V_0 + \cdots + V_n) = \frac{S}{n}, \tag{8}$$

where $S = V_0 + \cdots + V_n$ is the ($(n-1)$-dimensional) volume of the surface of K. Dividing (7) by (8), we obtain

$$\frac{1}{n} \ge \frac{V_0}{S} \cdot r_0 + \cdots + \frac{V_n}{S} \cdot r_n. \tag{9}$$

Note that the right hand of this inequality is a convex combination of $r_0, \ldots, r_n$ and therefore lower-bounded by

$$\min_{i=0,\ldots,n} r_i = r(K).$$

This completes the proof of Lemma 5.8.

We finally would like to mention that (8) implies that $V/S \le 1/n^2$. Quantity $1/n^2$ is precisely the volume-surface ratio of the regular $(n+1)$-simplex K_n. We have therefore accidentally proven that K_n achieves the highest volume-surface ratio among all $(n+1)$-simplices, which might be interesting in its own right.

Lemma 5.10 *Let $K = K_{n+1}$ be a $(n+1)$-simplex and $K_{n+1} \supset K_n \supset \cdots \supset K_0$ be a K-sequence of sub-simplexes with centers $z^{(n+1)}, z^{(n)}, \ldots, z^{(0)}$, respectively. Then, for all $k = n+1, \ldots, 0$, $\|z_k\|_2 \le R_{k,n}$, where $R_{k,n} = \sqrt{\frac{n-k+1}{kn}}$.*

Proof: We apply downward induction on k. Case $k = n+1$ is trivial and case $k = n$ is covered by Lemma 5.8. Let now k be fixed but arbitrary in the range from 0 to $n-1$. Remember that $K = K_{n+1}$ has center $z^{(n+1)} = \bar{0}$, K_k has center $z^{(k)}$, and K_n has center $z^{(n)}$. The situation is visualized in Figure 3(b), where $h_n = \|z^{(n)}\|_2$. Lemma 5.8 implies that $h_n \le 1/n$. In analogy to the proof of Lemma 5.4, we discuss the quantity $\tilde{R}_{k,n} = ||z^{(k)}||_2 = ||z^{(n+1)} - z^{(k)}||_2$ and obtain the recursion

$$\tilde{R}_{k,n}^2 = h_n^2 + (1 - h_n^2)\tilde{R}_{k,n-1}. \tag{10}$$

¿From the induction hypothesis, we conclude that $\tilde{R}_{k,n-1} \leq R_{k,n-1} \leq 1$. A comparison of the recursions 3 and 10 reveals that $\tilde{R}_{k,n} \leq R_{k,n}$, which concludes the proof. □

Corollary 5.11 *Let $K = K_{n+1}$ be a $(n+1)$-simplex and $K_{n+1} \supset K_n \supset \cdots \supset K_0$ be a K-sequence of sub-simplexes with centers $z^{(n+1)}, z^{(n)}, \ldots, z^{(0)}$, respectively. Then, for all $k = n, \ldots, 1$, the unit ball $B_* = B(\bar{0}, 1)$ is contained in $B(z_k, 1 + \sqrt{1/k})$.*

Proof: It suffices to show that $||z_k||_2 \leq \sqrt{1/k}$. This follows from Lemma 5.10 by observing that $\tilde{R}_{k,n}$ is increasing in n and approaches $\sqrt{1/k}$ when n approaches infinity. □

Let's finally glue things together. We argued in Subsection 5.1 that, without loss of generality, the input set $S \subset \Re^n$ contains $n+1$ points that are located on the sphere S_* and induce a non-marginal $(n+1)$-simplex K. Corollary 5.11 states that there exists a subset T of S of size k, namely the vertices of K_k, such that $B(z_k, 1 + \sqrt{1/k})$ contains B_*. Thus, if A_k outputs z_k, or even something superior to it, we are done. It suffices to show that z_k is among the candidate centers that A_k is inspecting. Since K is non-marginal and K_k belongs to a K-sequence, K_k is also non-marginal. This amounts in saying that the smallest ball containing T has center z_k. Thus, z_k is indeed one of the candidate centers of algorithm A_k. This concludes the proof of Theorem 5.1.

6 CONCLUSIONS

We briefly mention some extensions of our work that will be described in greater detail in the full paper:

- 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 that the cost-preserving reduction from the Densest Open Hemisphere Problem to the Densest (Open or Closed) Ball Problem is basically preserving the parameter μ. More precisely, the μ-relaxation of DOH is reduced to the $1 - \sqrt{1-\mu^2}$-relaxation of DOB or DCB. This holds also for functions μ, which typically approach zero when n approaches infinity. For μ being close to zero, term $1 - \sqrt{1-\mu^2}$ roughly equals μ. We call reductions of this type "margin-preserving" in the sequel.

- It follows that our family A_k of algorithms for the μ-relaxation of DCB can be combined with the polynomial cost- and margin-preserving reduction from DOH to DCB. The resulting algorithm finds a densest hemisphere given that there exists a densest hemisphere that yields a sufficiently large margin. This seems to be a nice alternative to perceptron-style algorithms.

- It can be shown that the $o(\sqrt{1/n})$-relaxation of DOH (and thus also of DOB or DCB) is still an NP-hard approximation problem. On the other hand, it is possibly to solve, in polynomial time, the $\Omega(\sqrt{1/\log n})$-relaxation of DCB (and thus also of DOH). We were not able to close the gap between these two bounds.

References

[1] Shai Ben-David, Nadav Eiron, and Philip Long. On the difficulty of approximately maximizing agreements. Submitted to this conference.

[2] 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.

[3] Shai Ben-David and Hans Ulrich Simon. Efficient learning of linear perceptrons, 2000. TR CS-2000-07, Computer Science Department, Technion.

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

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

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

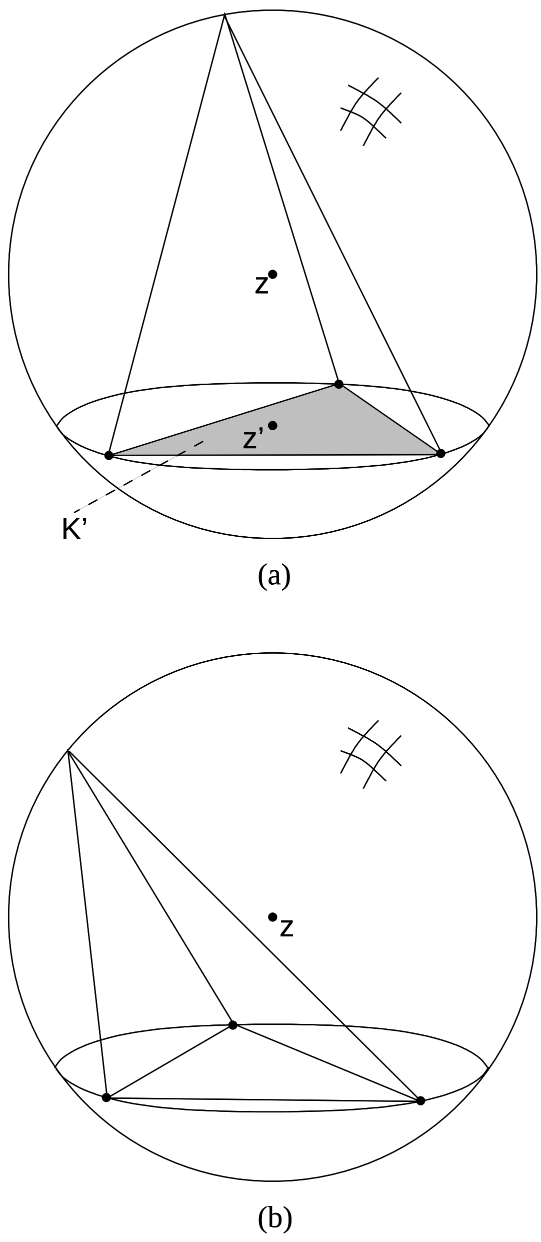

Figure 1: (a) A non-marginal 4-simplex K with its center z inside K, a face K' and the center z' of K'. (b) A marginal 4-simplex K with its center z outside K.

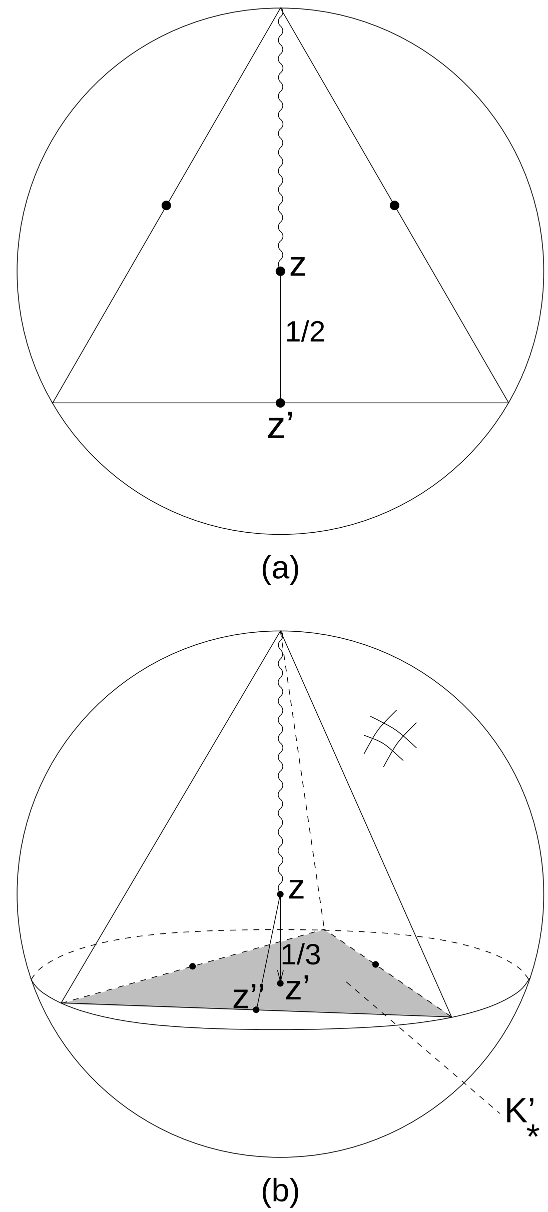

Figure 2: (a) The regular 3-simplex (augmented with some additional information), its center $z = \bar{0}$ and the center z' of one of its faces. (b) The regular 4-simplex (augmented with some additional information), its center $z = \bar{0}$, the center z' of one of its faces, and the center z'' of one of its 2-sub-simplexes.

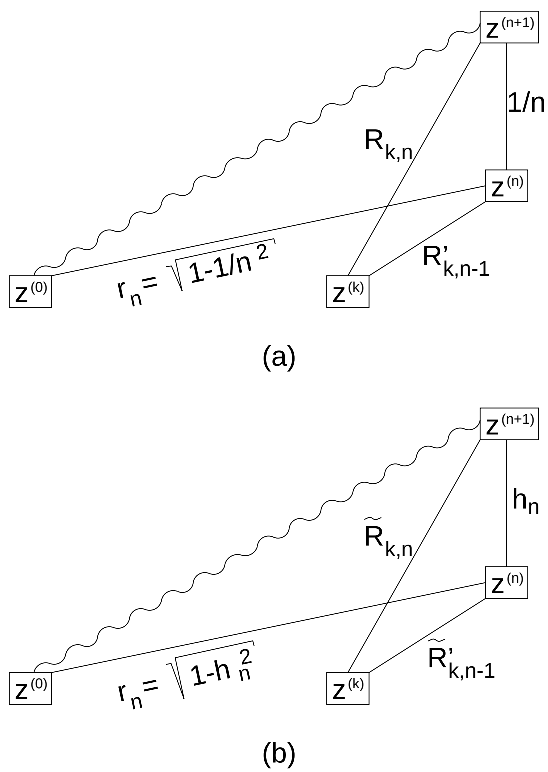

Figure 3: (a) Two triangles (with a right angle at $z^{(n)}$, respectively) induced by the center $z^{(n+1)}$ of the regular $(n+1)$-simplex K_* and the centers of some sub-simplexes of K_*. (b) The corresponding two triangles (with a right angle at $z^{(n)}$, respectively) in case of an arbitrary $(n+1)$-simplex K.

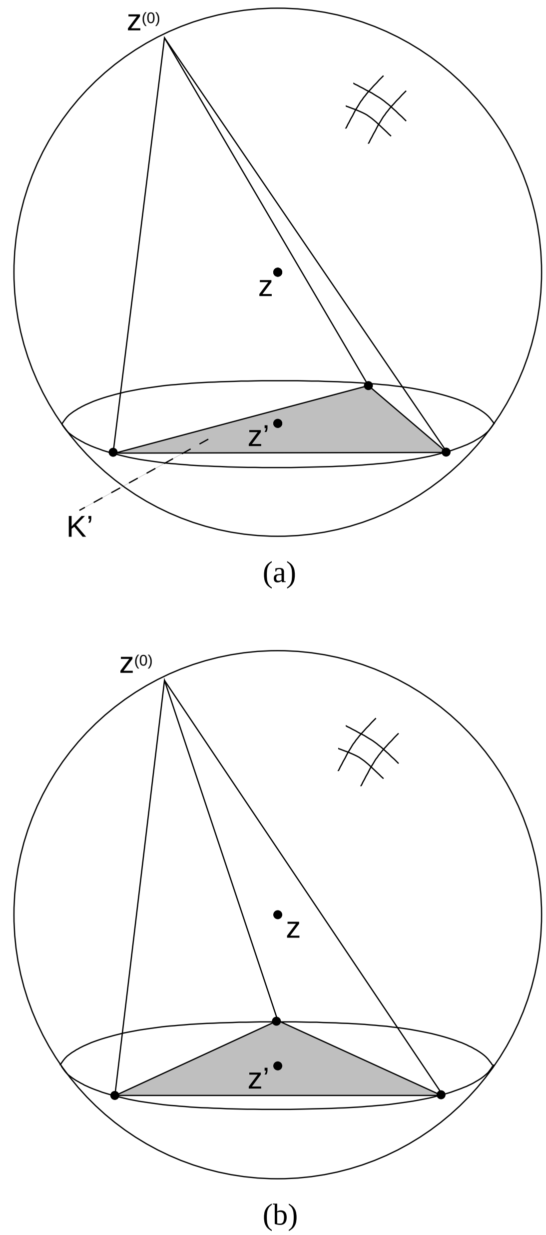

Figure 4: (a) A non-regular $(n+1)$-simplex K with a non-regular face K'. (b) The simplex that results from K by replacing K' by its regularization.

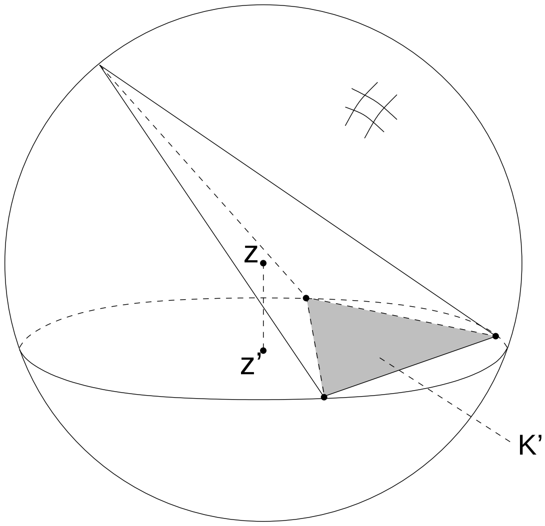

Figure 5: A non-marginal 4-simplex K with center z, a marginal face K' with center z'.

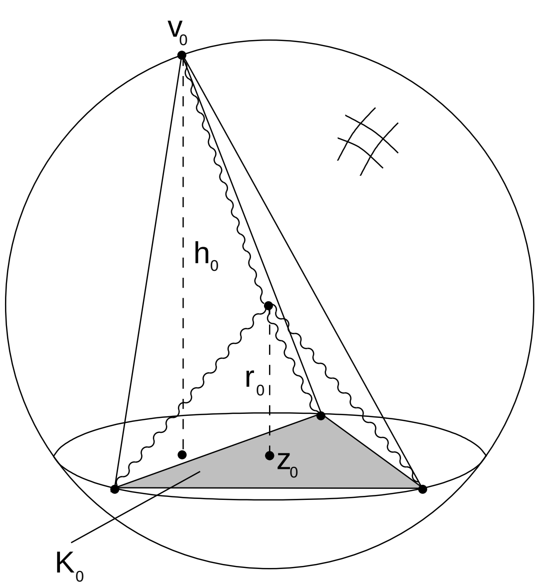

Figure 6: A 4-simplex illustrating the notations used in the proof of Lemma 5.8.