

New Upper Bounds on the Quality of the PCA Bounding Boxes in \mathbb{R}^2 and \mathbb{R}^3

Darko Dimitrov

Christian Knauer

Klaus Kriegel

Günter Rote

Institut für Informatik
Freie Universität Berlin
Takustrasse 9
D-14195 Berlin, Germany

darko/knauer/kriegel/rote@inf.fu-berlin.de

ABSTRACT

Principal component analysis (PCA) is commonly used to compute a bounding box of a point set in \mathbb{R}^d . The popularity of this heuristic lies in its speed, easy implementation and in the fact that usually, PCA bounding boxes quite well approximate the minimum-volume bounding boxes. Since there are examples of discrete points sets in the plane, showing that the worst case ratio of the volume of the PCA bounding box and the volume of the minimum-volume bounding box tends to infinity, we consider PCA bounding boxes for continuous sets, especially for the convex hull of a point set. Here, we contribute new upper bounds on the approximation factor of PCA bounding boxes of convex sets in \mathbb{R}^2 and \mathbb{R}^3 .

Categories and Subject Descriptors

F.2.2 [Nonnumerical Algorithms and Problems]: Geometrical problems and computations

General Terms

Algorithms, Theory

Keywords

Principal component analysis, Bounding boxes

1. INTRODUCTION

Substituting sets of points or complex geometric shapes with their bounding boxes is motivated by many applications. For example, in computer graphics, it is used to maintain hierarchical data structures for fast rendering of a scene or for collision detection. Additional applications include those in shape analysis and shape simplification, or

in statistics, for storing and performing range-search queries on a large database of samples.

Computing a minimum-area bounding box of a set of n points in \mathbb{R}^2 can be done in $O(n \log n)$ time, for example with the rotating caliper algorithm [13]. O'Rourke [10] presented a deterministic algorithm, a rotating caliper variant in \mathbb{R}^3 , for computing the minimum-volume bounding box of a set of n points in \mathbb{R}^3 . His algorithm requires $O(n^3)$ time and $O(n)$ space. Barequet and Har-Peled [2] have contributed two $(1+\epsilon)$ -approximation algorithms for computing the minimum-volume bounding box for point sets in \mathbb{R}^3 , both with nearly linear complexity. The running times of their algorithms are $O(n + 1/\epsilon^{4.5})$ and $O(n \log n + n/\epsilon^3)$, respectively.

Numerous heuristics have been proposed for computing a box which encloses a given set of points. The simplest heuristic is naturally to compute the axis-aligned bounding box of the point set. Two-dimensional variants of this heuristic include the well-known *R-tree*, the *packed R-tree* [11], the *R*-tree* [3], the *R+-tree* [12], etc.

A frequently used heuristic for computing a bounding box of a set of points is based on *principal component analysis*. The principal components of the point set define the axes of the bounding box. Once the axis directions are given, the dimension of the bounding box is easily found by the extreme values of the projection of the points on the corresponding axis. Two distinguished applications of this heuristic are the *OBB-tree* [5] and the *BOXTREE* [1], hierarchical bounding box structures, which support efficient collision detection and ray tracing. Computing a bounding box of a set of points in \mathbb{R}^2 and \mathbb{R}^3 by PCA is simple and requires linear time. To avoid the influence of the distribution of the point set on the directions of the PCs, a possible approach is to consider the convex hull, or the boundary of the convex hull $CH(P)$ of the point set P . Thus, the complexity of the algorithm increases to $O(n \log n)$. The popularity of this heuristic, besides its speed, lies in its easy implementation and in the fact that usually PCA bounding boxes are tight-fitting, c.f. see [8] for some experimental results.

Given a point set $P \subseteq \mathbb{R}^d$ we denote by $BB_{pca}(P)$ the PCA bounding box of P and by $BB_{opt}(P)$ the bounding box of P with smallest possible volume. The ratio of the two volumes $\lambda_d(P) = \text{Vol}(BB_{pca}(P))/\text{Vol}(BB_{opt}(P))$ de-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SCG'07, June 6–8, 2007, Gyeongju, South Korea.

Copyright 2007 ACM 978-1-59593-705-6/07/0006 ...\$5.00.

defines the approximation factor for P , and

$$\lambda_d = \sup \left\{ \lambda_d(P) \mid P \subseteq \mathbb{R}^d, \text{Vol}(CH(P)) > 0 \right\}$$

defines the general PCA approximation factor. To the best of our knowledge, the only known results about the quality of the PCA bounding boxes were given in [4], where lower bounds on λ_d for arbitrary dimension d , and an upper bound on λ_2 were presented. Here, we give a new upper bound on λ_2 , and the first upper bound on λ_3 .

The paper is organized as follows: In Section 2 we review the basics of principal component analysis and the known results about the quality of PCA bounding boxes. In particular, we introduce the continuous version of PCA, which results in a series of approximation factors $\lambda_{d,i}$, where i ranges from 0 to d and denotes the dimension of the faces of the convex hull that contribute to the continuous point set for which the principal components are computed. In Section 3 we give an upper bound on $\lambda_{2,2}$ and an upper bound on $\lambda_{3,3}$. We conclude with future work and open problems in Section 4.

2. PRINCIPAL COMPONENT ANALYSIS AND PCA BOUNDING BOXES

The central idea and motivation of PCA [7] (also known as the Karhunen-Loeve transform, or the Hotelling transform) is to reduce the dimensionality of a point set by identifying the most significant directions (principal components). Let $X = \{x_1, x_2, \dots, x_m\}$, where x_i is a d -dimensional vector, and $c = (c_1, c_2, \dots, c_d) \in \mathbb{R}^d$ be the center of gravity of X . For $1 \leq k \leq d$, we use x_{ik} to denote the k -th coordinate of the vector x_i . Given two vectors u and v , we use $\langle u, v \rangle$ to denote their inner product. For any unit vector $v \in \mathbb{R}^d$, the variance of X in direction v is

$$\text{var}(X, v) = \frac{1}{m} \sum_{i=1}^m \langle x_i - c, v \rangle^2. \quad (1)$$

The most significant direction corresponds to the unit vector v_1 such that $\text{var}(X, v_1)$ is maximum. In general, after identifying the j most significant directions $B_j = \{v_1, v_2, \dots, v_j\}$, the $(j+1)$ -th most significant direction corresponds to the unit vector v_{j+1} such that $\text{var}(X, v_{j+1})$ is maximum among all unit vectors perpendicular to v_1, v_2, \dots, v_j .

It can be verified that for any unit vector $v \in \mathbb{R}^d$,

$$\text{var}(X, v) = \langle Cv, v \rangle, \quad (2)$$

where C is the covariance matrix of X . C is a symmetric $d \times d$ matrix where the (i, j) -th component, c_{ij} , $1 \leq i, j \leq d$, is defined as

$$c_{ij} = \frac{1}{m} \sum_{k=1}^m (x_{ik} - c_i)(x_{jk} - c_j). \quad (3)$$

The procedure of finding the most significant directions, in the sense mentioned above, can be formulated as an eigenvalue problem. If $\lambda_1 > \lambda_2 > \dots > \lambda_d$ are the eigenvalues of C , then the unit eigenvector v_j for λ_j is the j -th most significant direction. All λ_j s are non-negative and $\lambda_j = \text{var}(X, v_j)$. Since the matrix C is symmetric positive definite, its eigenvectors are orthogonal. If the eigenvalues are not distinct, the eigenvectors are not unique. In this case, an orthogonal basis of eigenvectors is chosen arbitrary. However, we can

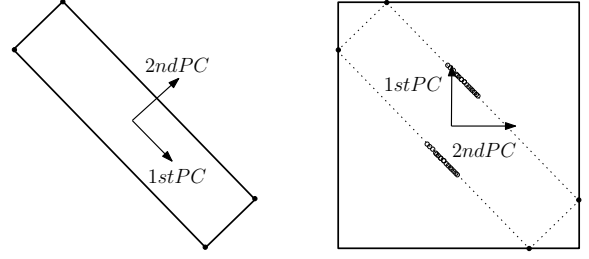


Figure 2: Four points and its PCA bounding-box (left). Dense collection of additional points significantly affect the orientation of the PCA bounding-box (right).

achieve distinct eigenvalues by a slight perturbation of the point set.

The following result summarizes the above background knowledge on PCA. For any set S of orthogonal unit vectors in \mathbb{R}^d , we use $\text{var}(X, S)$ to denote $\sum_{v \in S} \text{var}(X, v)$.

LEMMA 1. For $1 \leq j \leq d$, let λ_j be the j -th largest eigenvalue of C and let v_j denote the unit eigenvector for λ_j . Let $B_j = \{v_1, v_2, \dots, v_j\}$, $\text{sp}(B_j)$ be the linear subspace spanned by B_j , and $\text{sp}(B_j)^\perp$ be the orthogonal complement of $\text{sp}(B_j)$. Then $\lambda_1 = \max\{\text{var}(X, v) : v \in \mathbb{R}^d, \|v\| = 1\}$, and for any $2 \leq j \leq d$,

$$i) \lambda_j = \max\{\text{var}(X, v) : v \in \text{sp}(B_{j-1})^\perp, \|v\| = 1\}.$$

$$ii) \lambda_j = \min\{\text{var}(X, v) : v \in \text{sp}(B_j), \|v\| = 1\}.$$

$$iii) \text{var}(X, B_j) \geq \text{var}(X, S) \text{ for any set } S \text{ of } j \text{ orthogonal unit vectors.}$$

Since bounding boxes of a point set P (with respect to any orthogonal coordinate system) depend only on the convex hull of $CH(P)$, the construction of the covariance matrix should be based only on $CH(P)$ and not on the distribution of the points inside. Using the vertices, i.e., the 0-dimensional faces of $CH(P)$ to define the covariance matrix C we obtain a bounding box $BB_{pca(d,0)}(P)$. We denote by $\lambda_{d,0}(P)$ the approximation factor for the given point set P and by

$$\lambda_{d,0} = \sup \left\{ \lambda_{d,0}(P) \mid P \subseteq \mathbb{R}^d, \text{Vol}(CH(P)) > 0 \right\}$$

the approximation factor in general. The example in Figure 2 shows that $\lambda_{2,0}(P)$ can be arbitrarily large if the convex hull is nearly a thin rectangle, with a lot of additional vertices in the middle of the two long sides. Since this construction can be lifted into higher dimensions we obtain a first general lower bound.

PROPOSITION 1. $\lambda_{d,0} = \infty$ for any $d \geq 2$.

To overcome this problem, one can apply a continuous version of PCA taking into account (the dense set of) all points on the boundary of $CH(P)$, or even all points in $CH(P)$. In this approach X is a continuous set of d -dimensional vectors and the coefficients of the covariance matrix are defined by integrals instead of finite sums. The computation of the coefficients of the covariance matrix in the continuous case can be done also in linear time, thus, the overall complexity remains the same as in the discrete case. Note that for

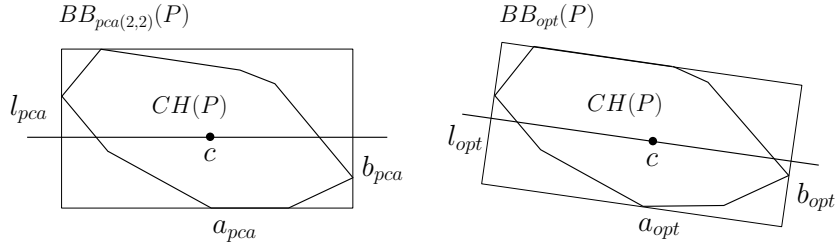


Figure 1: A convex hull of the point set P , its PCA bounding box and its optimal bounding box.

for $d = 1$ the above problem is trivial, because the PCA bounding box is always optimal, i.e., $\lambda_{1,0}$ and $\lambda_{1,1}$ are 1.

Variants of the continuous PCA, applied on triangulated surfaces of 3D objects, were presented by Gottschalk et al. [5], Lahanas et al. [8] and Vranić et al. [14]. For point sets P in \mathbb{R}^2 we are especially interested in the cases when X represents the boundary of $CH(P)$, or all points in $CH(P)$. Since the first case corresponds to the 1-dimensional faces of $CH(P)$ and the second case to the 2-dimensional face of $CH(P)$, the generalization to dimension $d > 2$ leads to a series of $d - 1$ continuous PCA versions. For a point set $P \in \mathbb{R}^d$, $C(P, i)$ denotes the covariance matrix defined by the points on the i -dimensional faces of $CH(P)$, and $BB_{pca(d,i)}(P)$, denotes the corresponding bounding box. The approximation factors $\lambda_{d,i}(P)$ and $\lambda_{d,i}$ are defined as

$$\lambda_{d,i}(P) = \frac{\text{Vol}(BB_{pca(d,i)}(P))}{\text{Vol}(BB_{opt}(P))}, \quad \text{and}$$

$$\lambda_{d,i} = \sup \{ \lambda_{d,i}(P) \mid P \subseteq \mathbb{R}^d, \text{Vol}(CH(P)) > 0 \}.$$

In what follows, we give a brief overview of the results from [4], which to the best of our knowledge are the only known bounds on the quality of the PCA bounding boxes. First, we present an extension of Proposition 1, which indicates that for a given d , there remain only two interesting cases: the factor $\lambda_{d,d-1}$ corresponding to the boundary of the convex hull, and the factor $\lambda_{d,d}$ corresponding to the full convex hull.

PROPOSITION 2. $\lambda_{d,i} = \infty$ for any $d \geq 2$ and any $0 \leq i < d - 1$.

The following nontrivial lower bounds are based on the relation between the symmetry of a point set and its principal components [4, Lemma 4].

THEOREM 1. $\lambda_{3,2} \geq 4$ and $\lambda_{3,3} \geq 4$.

THEOREM 2. If d is a power of two, then $\lambda_{d,d-1} \geq \sqrt{d}^d$ and $\lambda_{d,d} \geq \sqrt{d}^d$.

In [4] also the first nontrivial upper bound on $\lambda_{2,1}$ is given.

THEOREM 3. The PCA bounding box of a point set P in \mathbb{R}^2 computed over the boundary of $CH(P)$ has a guaranteed approximation factor $\lambda_{2,1} \leq 2.737$.

Although this result concerns a continuous PCA version, the proof is mainly based on arguments from discrete geometry. In contrast to that, the upper bound proofs presented in this paper essentially make use of integral calculus. In what follows we present the first upper bounds on $\lambda_{2,2}$ and $\lambda_{3,3}$.

3. NEW UPPER BOUNDS

3.1 An upper bound for $\lambda_{2,2}$

Given a point set $P \subseteq \mathbb{R}^2$ and an arbitrary bounding box $BB(P)$, we will denote the two side lengths of $BB(P)$ by a and b , where $a \geq b$. We are interested in the side lengths $a_{opt}(P) \geq b_{opt}(P)$ and $a_{pca}(P) \geq b_{pca}(P)$ of $BB_{opt}(P)$ and $BB_{pca(2,2)}(P)$, see Figure 1. The parameters $\alpha = \alpha(P) = a_{pca}(P)/a_{opt}(P)$ and $\beta = \beta(P) = b_{pca}(P)/b_{opt}(P)$ denote the ratios between the corresponding side lengths, so that $\lambda_{2,2}(P) = \alpha(P) \cdot \beta(P)$. If the relation to P is clear, we will omit the reference to P in the notations introduced above.

Since the side lengths of any bounding box are bounded by the diameter of P , we can observe that in general $b_{pca}(P) \leq a_{pca}(P) \leq \text{diam}(P) \leq \sqrt{2}a_{opt}(P)$, and in the special case when the optimal bounding box is a square $\lambda_{2,2}(P) \leq 2$. This observation can be generalized, introducing an additional parameter $\eta(P) = a_{opt}(P)/b_{opt}(P)$.

LEMMA 2. $\lambda_{2,2}(P) \leq \eta + \frac{1}{\eta}$ for any point set P with aspect ratio $\eta(P) = \eta$.

PROOF. For both a_{pca} and b_{pca} , we have the upper bound $\text{diam}(P) \leq \sqrt{a_{opt}^2 + b_{opt}^2} = a_{opt} \sqrt{1 + \frac{1}{\eta^2}}$. Thus, $\alpha\beta = \frac{a_{pca}b_{pca}}{a_{opt}b_{opt}} \leq \frac{(a_{opt} \sqrt{1 + \frac{1}{\eta^2}})^2}{a_{opt}b_{opt}} = \frac{a_{opt}}{b_{opt}} (1 + \frac{1}{\eta^2})$. Replacing a_{opt} by $\eta \cdot b_{opt} = \eta (1 + \frac{1}{\eta^2}) = \eta + \frac{1}{\eta}$. \square

Unfortunately, this parametrized upper bound tends to infinity for $\eta \rightarrow \infty$. Therefore, we are going to derive another upper bound that is better for large values of η . We derive such a bound by finding a constant that bounds β from above. In this process we will make essential use of the properties of $BB_{pca(2,2)}(P)$. We denote by $d^2(CH(P), l)$ the integral of the squared distances of the points on $CH(P)$ to a line l , i.e., $d^2(CH(P), l) = \int_{s \in CH(P)} d^2(s, l) ds$. Let l_{pca} be the line going through the center of gravity, parallel to the longer side of $BB_{pca(2,2)}(P)$, and l_{opt} be the line going through the center of gravity, parallel to the longer side of $BB_{opt}(P)$ (see Figure 1). By Lemma 1, part ii) l_{pca} is the best fitting line of P and therefore

$$d^2(CH(P), l_{pca}) \leq d^2(CH(P), l_{opt}). \quad (4)$$

We obtain an estimate for β by determining a lower bound on $d^2(CH(P), l_{pca})$ that depends on b_{pca} , and an upper bound on $d^2(CH(P), l_{opt})$ that depends on b_{opt} . Having an arbitrary bounding box of $CH(P)$ (with side lengths a and b , $a \geq b$) the area of $CH(P)$ can be expressed as

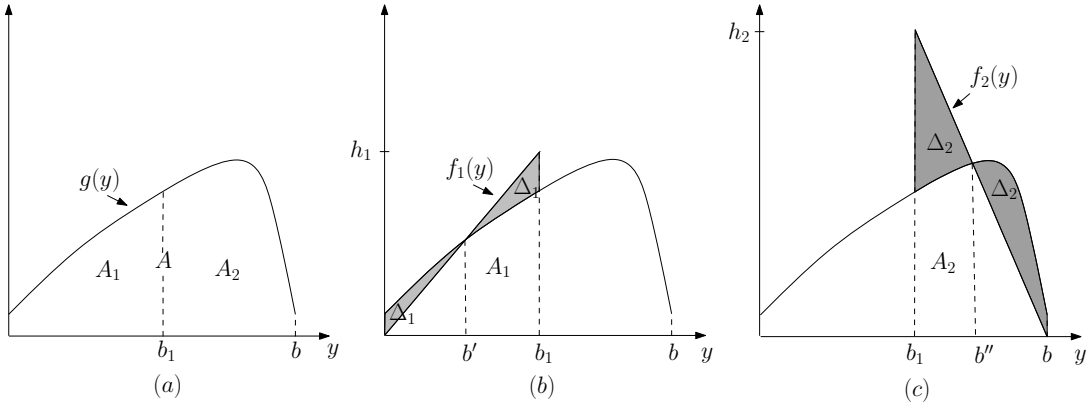


Figure 3: Construction of the lower bound for $d^2(CH(P), l_{b_1})$.

$$A = A(CH(P)) = \int_0^b \int_0^a \chi_{CH(P)}(x, y) dx dy = \int_0^b g(y) dy,$$

where $\chi_{CH(P)}(x, y)$ is the characteristic function of $CH(P)$ defined as

$$\chi_{CH(P)}(x, y) = \begin{cases} 1 & (x, y) \in CH(P) \\ 0 & (x, y) \notin CH(P), \end{cases}$$

and $g(y) = \int_0^a \chi_{CH(P)}(x, y) dx$ is the length of the intersection of $CH(P)$ with a horizontal line at height y . In the following we call $g(y)$ the density function of $CH(P)$ for computing the area with the integral $\int_0^b g(y) dy$. Note that $g(y)$ is continuous and convex in the interval $[0, b]$ (see Figure 3 (a) for an illustration). Let b_1 denote the y -coordinate of the center of gravity of $CH(P)$. The line l_{b_1} ($y = b_1$) divides the area of the $CH(P)$ into A_1 and A_2 .

Theorem 5, which is derived from the generalized first mean value theorem of integral calculus (Theorem 4), is our central technical tool in derivation of the lower and the upper bound on $d^2(CH(P), l_{b_1})$.

THEOREM 4. (Generalized first mean value theorem of integral calculus) *If $h(x)$ and $g(x)$ are continuous functions in the interval $[a, b]$, and if $g(x)$ does not change its sign in the interval, then there is a $\xi \in (a, b)$ such that*

$$\int_a^b h(x)g(x)dx = h(\xi) \int_a^b g(x)dx.$$

THEOREM 5. *Let $f(x)$ and $g(x)$ be positive continuous functions on the interval $[a, b]$ with $\int_a^b f(x)dx = \int_a^b g(x)dx$, and assume that there is some $c \in [a, b]$ such that $f(x) \leq g(x)$, for all $x \leq c$ and $f(x) \geq g(x)$, for all $x \geq c$. Then*

$$\int_a^b (x-b)^2 f(x)dx \leq \int_a^b (x-b)^2 g(x)dx \quad \text{and}$$

$$\int_a^b (x-a)^2 f(x)dx \geq \int_a^b (x-a)^2 g(x)dx.$$

PROOF. We start from the assumptions $\int_a^b f(x)dx = \int_a^b g(x)dx$ and $f(x) \leq g(x)$ for all $x \leq c$ and $f(x) \geq g(x)$ for all $x \geq c$.

Thus,

$$\int_a^c (g(x) - f(x))dx = \int_c^b (f(x) - g(x))dx = \Delta \quad (5)$$

and the integrands on both sides are nonnegative. Applying Theorem 4 to the following integrals we obtain

$$\begin{aligned} \int_a^c (x-b)^2 (g(x) - f(x))dx &= (\xi_1 - b)^2 \int_a^c (g(x) - f(x))dx \\ &= (\xi_1 - b)^2 \Delta, \end{aligned}$$

and

$$\begin{aligned} \int_c^b (x-b)^2 (f(x) - g(x))dx &= (\xi_2 - b)^2 \int_c^b (f(x) - g(x))dx \\ &= (\xi_2 - b)^2 \Delta, \end{aligned}$$

for some $\xi_1 \in [a, c]$ and $\xi_2 \in [c, b]$. Therefore

$$\begin{aligned} \int_a^c (x-b)^2 (g(x) - f(x))dx &= (\xi_1 - b)^2 \Delta \geq (\xi_2 - b)^2 \Delta \\ &= \int_c^b (x-b)^2 (f(x) - g(x))dx. \end{aligned}$$

It follows that

$$\begin{aligned} \int_a^b (x-b)^2 (g(x) - f(x))dx &= \\ \int_a^c (x-b)^2 (g(x) - f(x))dx - \int_c^b (x-b)^2 (f(x) - g(x))dx &\geq 0 \end{aligned}$$

what proves the first claim

$$\int_a^b (x-b)^2 f(x)dx \leq \int_a^b (x-b)^2 g(x)dx.$$

The proof of the second claim follows from symmetry. \square

The following theorem was discovered independently by Grünbaum [6] and Hammer (unpublished manuscript), and later rediscovered by Mityagin [9]. We use it to prove a lower and an upper bound of the variance $d^2(CH(P), l_{b_1})$.

THEOREM 6 (GRÜNBAUM-HAMMER-MITYAGIN). *Let K be a compact convex set in \mathbb{R}^d with nonempty interior and centroid μ . Assume that the d -dimensional volume of K is one, that is, $\text{Vol}_d(K) = 1$. Let H be any $(d-1)$ -dimensional plane passing through μ with corresponding half-spaces H^+ and H^- . Then,*

$$\min\{\text{Vol}_d(K \cap H^+), \text{Vol}_d(K \cap H^-)\} \geq \left(\frac{d}{d+1}\right)^d$$

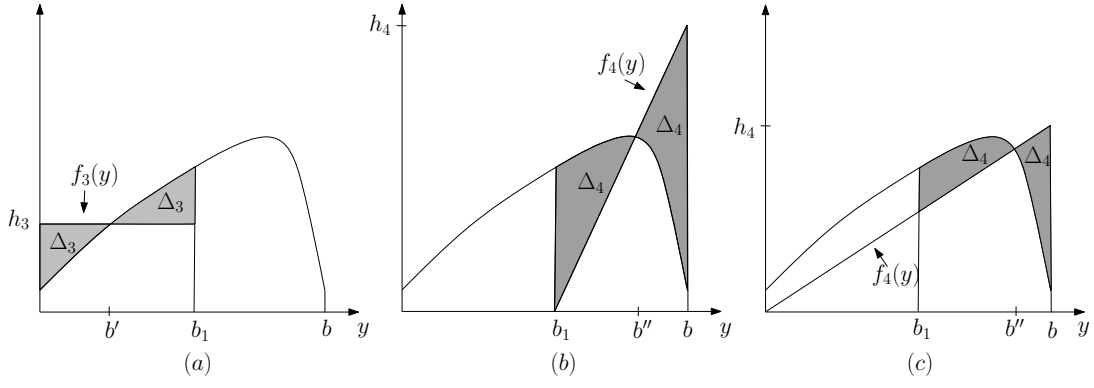


Figure 4: Construction of the upper bound for $d^2(CH(P), l_{b_1})$.

Moreover, the bound $(\frac{d}{d+1})^d$ is best possible.

LEMMA 3. The variance $d^2(CH(P), l_{b_1})$ is bounded from below by $\frac{10}{243}Ab^2$.

PROOF. We split the integral $\int_0^b (y - b_1)^2 g(y) dy$ at b_1 , and prove lower bounds for both parts in the following way: For the left part consider the linear function $f_1(y) = \frac{h_1}{b_1}y$ such that $\int_0^{b_1} f_1(y) dy = \int_0^{b_1} g(y) dy = A_1$ (see Figure 3 (b) for an illustration). From $\int_0^{b_1} f_1(y) dy = A_1$, it follows that $f_1(y) = \frac{2A_1 y}{b_1^2}$. Since $g(y)$ is convex, $g(y)$ and $f_1(y)$ intersect only once, at point $b' \in (0, b_1)$. By Theorem 5, we have

$$\begin{aligned} \int_0^{b_1} (y - b_1)^2 g(y) dy &\geq \int_0^{b_1} (y - b_1)^2 f_1(y) dy = \\ \int_0^{b_1} (y - b)^2 \frac{2A_1}{b_1^2} dy &= \frac{A_1 b_1^2}{6}. \end{aligned} \quad (6)$$

Analogously, for the right part consider the linear function $f_2(y) = \frac{h_2}{b_1 - b}(y - b) = \frac{h_2}{-b_2}(y - b)$ such that $\int_{b_1}^b f_2(y) dy = \int_{b_1}^b g(y) dy = A_2$ (see Figure 3 (c) for an illustration). From $\int_{b_1}^b f_2(y) dy = A_2$, it follows that $f_2(y) = \frac{2A_2}{b_2^2}(y - b)$. Since $g(y)$ is convex, $g(y)$ and $f_2(y)$ intersect only once, at point $b' \in (b_1, b)$. By Theorem 5, we have that

$$\begin{aligned} \int_{b_1}^b (y - b_1)^2 g(y) dy &\geq \int_{b_1}^b (y - b_1)^2 f_2(y) dy = \\ \int_{b_1}^b (y - b_1)^2 \frac{2A_2}{(b - b_1)^2} (y - b_1) dy &= \frac{A_2 b_2^2}{6}. \end{aligned} \quad (7)$$

From (6) and (7) we obtain that

$$\begin{aligned} d^2(CH(P), l_{b_1}) &= \int_0^{b_1} (y - b_1)^2 g(y) dy + \\ \int_{b_1}^b (y - b_1)^2 g(y) dy &\geq \frac{A_1 b_1^2}{6} + \frac{A_2 b_2^2}{6}. \end{aligned}$$

From the Grünbaum-Hammer-Mityagin theorem (see Appendix), we know that $A_1, A_2 \in [\frac{4}{9}A, \frac{5}{9}A]$. Also, we know that $b_1, b_2 \in [\frac{1}{3}b, \frac{2}{3}b]$. It is not hard to show that, under these constraints, the expression $\frac{A_1 b_1^2}{6} + \frac{A_2 b_2^2}{6}$ achieves its minimum of $\frac{10}{243}Ab^2$ for $A_1 = \frac{4}{9}A, b_1 = \frac{5}{9}b$ or $A_1 = \frac{5}{9}A, b_1 = \frac{4}{9}b$. \square

LEMMA 4. The variance $d^2(CH(P), l_{b_1})$ is bounded from above by $\frac{29}{243}Ab^2$.

PROOF. Without loss of generality, we can assume that $g(y)$ has its maximum in $[b_1, b]$. We split the integral $\int_0^b (y - b_1)^2 g(y) dy$ at b_1 , and prove upper bounds for both parts in the following way. For the left part consider a linear function $f_3(y) = h_3$ such that $\int_0^{b_1} f_3(y) dy = \int_0^{b_1} g(y) dy = A_1$ (see Figure 4 (a) for an illustration). This implies that $f_3(y) = \frac{A_1}{b_1}$, and since $g(y)$ is convex, $g(y)$ and $f_3(y)$ intersect only once, at point $b' \in (b_1, b)$. By Theorem 5, we have

$$\begin{aligned} \int_0^{b_1} (y - b_1)^2 g(y) dy &\leq \int_0^{b_1} (y - b_1)^2 f_3(y) dy = \\ \int_0^{b_1} (y - b_1)^2 \frac{A_1}{b_1} dy &= \frac{A_1 b_1^2}{3}. \end{aligned} \quad (8)$$

Now, we are looking for an appropriate function $f_4(y)$ to derive an upper bound of the second part of the integral $\int_0^b (y - b_1)^2 g(y) dy$. Note that both functions $f_3(y)$ and $f_4(y)$, in general can not be of the type $f(y) = const$, because it can happen that $f_4(y)$ intersects $g(y)$ twice, and we can not apply Theorem 5. Thus, for the left part we consider a linear function $f_4(y) = \frac{h_2}{b}y$ such that $\int_{b_1}^b f_4(y) dy = \int_{b_1}^b g(y) dy = A_2$ (see Figure 4 (c) for an illustration). $\int_{b_1}^b f_4(y) dy = A_2$ implies that $f_4(y) = \frac{2A_2 b_1}{b_2(b_1 + b)}y$, and since $g(y)$ is convex, $g(y)$ and $f_4(y)$ intersect only once, at point $b'' \in (b_1, b)$. By Theorem 5, we have

$$\begin{aligned} \int_{b_1}^b (y - b_1)^2 g(y) dy &\geq \int_{b_1}^b (y - b_1)^2 f_4(y) dy = \\ \int_{b_1}^b (y - b_1)^2 \frac{2A_2 b_1}{b_2(b_1 + b)} y dy &= \frac{A_2 b_2^2}{b_1 + b} \left(\frac{b_1}{4} + \frac{b_2}{4} \right). \end{aligned} \quad (9)$$

From (8) and (9) we obtain

$$\begin{aligned} d^2(\mathcal{P}, l_{b_1}) &= \int_0^{b_1} (y - b_1)^2 g(y) dy + \int_{b_1}^b (y - b_1)^2 g(y) dy \\ &\leq \frac{A_1 b_1^2}{3} + \frac{A_2 b_2^2}{b_1 + b} \left(\frac{b_1 + b_2}{4} \right). \end{aligned}$$

From the Grünbaum-Hammer-Mityagin theorem, we know that $A_1, A_2 \in [\frac{4}{9}A, \frac{5}{9}A]$. Also, we know that $b_1, b_2 \in [\frac{1}{3}b, \frac{2}{3}b]$. It is not hard to show that, under these constraints, the expression $\frac{A_1 b_1^2}{3} + \frac{A_2 b_2^2}{b_1 + b} \left(\frac{b_1 + b_2}{4} \right)$ achieves its minimum of $\frac{29}{243}Ab^2$ for $A_1 = \frac{4}{9}A, b_1 = \frac{1}{3}b$. \square

We remark that in Lemma 4 we can use the function $f_4(y) = \frac{h_4}{b_2}(y - b_1)$ instead of $f_4(y) = \frac{h_2}{b}y$ (see Figure 4 (b) for an

illustration), but that will give us bigger upper bound for $d^2(CH(P), l_{b_1})$, namely $\frac{34}{243}Ab^2$.

Now, we are ready to derive an alternative parametrized upper bound on $\lambda_{2,2}(P)$ which is better than the bound from Lemma 2 for big values of η .

LEMMA 5. $\lambda_{2,2}(P) \leq \sqrt{2.9 \left(1 + \frac{1}{\eta^2}\right)}$ for any point set P with aspect ratio $\eta(P) = \eta$.

PROOF. Applying Lemma 3 and Lemma 4 in (4) we obtain

$$\frac{10}{243}Ab_{pca}^2 \leq d^2(P, l_{pca}) \leq d^2(P, l_{opt}) \leq \frac{29}{243}Ab_{opt}^2. \quad (10)$$

From (10) it follows that $\beta = \frac{b_{pca}}{b_{opt}} \leq \sqrt{2.9}$. We have for a_{pca} the upper bound $diam(P) \leq \sqrt{a_{opt}^2 + b_{opt}^2} = a_{opt}\sqrt{1 + \frac{1}{\eta^2}}$.

From this, it follows that $\alpha \leq \sqrt{1 + \frac{1}{\eta^2}}$. Putting this together, we obtain $\alpha\beta \leq \sqrt{2.9 \left(1 + \frac{1}{\eta^2}\right)}$. \square

THEOREM 7. The PCA bounding box of a point set P in \mathbb{R}^2 computed over $CH(P)$ has a guaranteed approximation factor $\lambda_{2,2} \leq 2.104$.

PROOF. The theorem follows from the combination of the two parametrized bounds from Lemma 2 and Lemma 5:

$$\lambda_{2,2} \leq \sup_{\eta \geq 1} \left\{ \min \left(\eta + \frac{1}{\eta}, \sqrt{2.9 \left(1 + \frac{1}{\eta^2}\right)} \right) \right\}.$$

It is easy to check that the supremum $s \approx 2.1038$ is obtained for $\eta \approx 1.3784$. \square

3.2 An upper bound for $\lambda_{3,3}$

Some of the techniques used here are similar to those used in Subsection 3.1 where we derive an upper bound on $\lambda_{2,2}$. One essential difference is that for the upper bound for $\lambda_{3,3}$, we additionally need a bound for the ratio of the middle sides of $BB_{pca(3,3)}(P)$ and $BB_{opt}(P)$, which we derive from the relation in Lemma 9.

Given a point set $P \subseteq \mathbb{R}^3$ and an arbitrary bounding box $BB(P)$, we will denote the three side lengths of $BB(P)$ by a, b and c , where $a \geq b \geq c$. We are interested in the side lengths $a_{opt} \geq b_{opt} \geq c_{opt}$ and $a_{pca} \geq b_{pca} \geq c_{pca}$ of $BB_{opt}(P)$ and $BB_{pca(3,3)}(P)$. The parameters $\alpha = \alpha(P) = a_{pca}/a_{opt}$, $\beta = \beta(P) = b_{pca}/b_{opt}$ and $\gamma = \gamma(P) = c_{pca}/c_{opt}$ denote the ratios between the corresponding side lengths. Hence, we have $\lambda_{3,3}(P) = \alpha \cdot \beta \cdot \gamma$.

Since the side lengths of any bounding box are bounded by the diameter of P , we can observe that in general $c_{pca} \leq b_{pca} \leq a_{pca} \leq diam(P) \leq \sqrt{3}a_{opt}$, and in the special case when the optimal bounding box is a cube $\lambda_{3,3}(P) \leq 3\sqrt{3}$. This observation can be generalized, introducing two additional parameters $\eta(P) = a_{opt}/b_{opt}$ and $\theta(P) = a_{opt}/c_{opt}$.

LEMMA 6. $\lambda_{3,3}(P) \leq \eta\theta \left(1 + \frac{1}{\eta^2} + \frac{1}{\theta^2}\right)^{\frac{3}{2}}$ for any point set P with aspect ratios $\eta(P) = \eta$ and $\theta(P) = \theta$.

PROOF. We have for a_{pca} , b_{pca} and c_{pca} the upper bound $diam(P) \leq \sqrt{a_{opt}^2 + b_{opt}^2 + c_{opt}^2} = a_{opt}\sqrt{1 + \frac{1}{\eta^2} + \frac{1}{\theta^2}}$. Thus,

$$\alpha\beta\gamma \leq \frac{a_{pca}b_{pca}c_{pca}}{a_{opt}b_{opt}c_{opt}} \leq \frac{a_{opt}^3 \left(1 + \frac{1}{\eta^2}\right)^{\frac{3}{2}}}{a_{opt}b_{opt}c_{opt}}. \quad \text{Replacing } a_{opt} \text{ in}$$

the nominator once by ηb_{opt} and once by θc_{opt} we obtain $\lambda_{3,3}(P) \leq \eta\theta \left(1 + \frac{1}{\eta^2} + \frac{1}{\theta^2}\right)^{\frac{3}{2}}$. \square

Unfortunately, this parametrized upper bound tends to infinity for $\eta \rightarrow \infty$ or $\theta \rightarrow \infty$. Therefore we are going to derive another upper bound that is better for large values of η and θ . We derive such a bound by finding constants that bound β and γ from above. In this process we will make essential use of the properties of $BB_{pca(3,3)}(P)$. We denote by $d^2(CH(P), H)$ the integral of the squared distances of the points on $CH(P)$ to a plane H , i.e., $d^2(CH(P), H) = \int_{s \in CH(P)} d^2(s, H) ds$. Let H_{pca} be the plane going through the center of gravity, parallel to the side $a_{pca} \times b_{pca}$ of $BB_{pca(3,3)}(P)$, and H_{opt} be the bisector of $BB_{opt}(P)$ parallel to the side $a_{opt} \times b_{opt}$. By Lemma 1, part ii) H_{pca} is the best fitting plane of P and therefore

$$d^2(CH(P), H_{pca}) \leq d^2(CH(P), H_{opt}). \quad (11)$$

We obtain an estimation for β by determining a lower bound on $d^2(CH(P), H_{pca})$ that depends on b_{pca} , and an upper bound on $d^2(CH(P), H_{opt})$ that depends on b_{opt} . Having an arbitrary bounding box of $CH(P)$ (with side lengths a , b , and c , $a \geq b \geq c$) the volume of $CH(P)$ can be expressed as

$$V = V(CH(P)) = \int_0^c \int_0^b \int_0^a \chi_{CH(P)}(x, y, z) dx dy dz = \int_0^c g(z) dz,$$

where $\chi_{CH(P)}(x, y, z)$ is the characteristic function of $CH(P)$ defined as

$$\chi_{CH(P)}(x, y, z) = \begin{cases} 1 & (x, y, z) \in CH(P) \\ 0 & (x, y, z) \notin CH(P), \end{cases}$$

and $g(z) = \int_0^b \int_0^a \chi_{CH(P)}(x, y, z) dx dy$ is the area of the intersection of $CH(P)$ with the horizontal plane at height z . As before we call $g(z)$ the density function of $CH(P)$. Let c_1 denote the z -coordinate of the center of gravity of $CH(P)$. The line l_{c_1} ($y = c_1$) divides the volume of $CH(P)$ into V_1 and V_2 (see Figure 5 (1) for an illustration).

Note that $g(z)$ is continuous, but in general not convex in the interval $[0, b]$. Therefore, we can not use linear functions to derive a lower and an upper bound of the function $d^2(CH(P), H_{ab})$, as we did in Subsection 3.1, because a linear function can intersect $g(z)$ more than once, and we can not apply Theorem 5. Instead of linear functions, we use quadratic functions.

PROPOSITION 3. Let $g(z)$ be the density function of $CH(P)$ defined as above, and let $f(z) = kz^2$ be the parabola such that $\int_0^{c_1} f(z) dz = \int_0^{c_1} g(z) dz$. Then, $\exists c_0 \in [0, c_1]$ such that $f(z) \leq g(z)$ for all $z \leq c_0$ and $f(z) \geq g(z)$ for all $z \geq c_0$.

PROOF. We give a constructive proof. Let $c_0 := \inf \{ d \mid \forall z \in [d, c_1] g(z) \leq f(z) \}$. If $c_0 = 0$, then $f(z) = g(z)$, and the proposition holds. If $c_0 > 0$, then consider the polygon which is the intersection of $CH(P)$ with the plane $z = c_0$. We fix a point p_0 in $CH(P)$ with z -coordinate 0 and construct a pyramid Q by extending all rays from p_0 through the polygon up to the plane $z = c_1$ (see Figure 5 for an illustration). Since, $f(c_0) = g(c_0)$ the quadratic function $f(z)$ is the density function of Q . Therefore, since the part of Q below c_0 is completely included in the $CH(P)$, we can

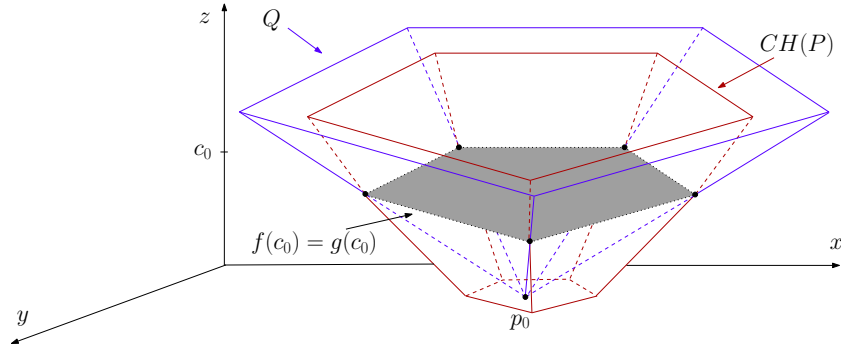


Figure 5: Construction of the intersection of $f(z)$ and $g(z)$.

conclude that $f(z) \leq g(z)$ for all $z \leq c_0$. On the other side, $f(z) \geq g(z)$ for all $z \geq c_0$ by the definition of c_0 . \square

Now, we present a lower and an upper bound on the variance $d^2(CH(P), H_{ab})$, from which we can derive a bound on $\gamma = \frac{c_{pca}}{c_{opt}}$.

LEMMA 7. *The variance $d^2(CH(P), H_{ab})$ is bounded from below by $\frac{7}{256}Vc^2$.*

PROOF. We split the integral $\int_0^c (z - c_1)^2 g(z) dz$ at c_1 , and prove upper bounds for both parts in the following way: For the left part consider the parabola $f_1(z) = \frac{h_1}{c_1^2} z^2$ such that $\int_0^{c_1} f_1(z) dz = \int_0^{c_1} g(z) dz = V_1$ (see Figure 6 (b) for an illustration). From $\int_0^{c_1} f_1(z) dz = V_1$ we have that $f_1(z) = \frac{3V_1}{c_1^3} z^2$. Since $f_1(z)$ and $g(z)$ define the same volume on the interval $[0, c_1]$, they must intersect, and by Proposition 3 we know that if $f_1(z) \neq g(z)$, then they can intersect only once, at a point $c' \in (0, c_1)$. Under these conditions, we can apply Theorem 5, and obtain

$$\begin{aligned} \int_0^{c_1} (z - c_1)^2 g(z) dz &\geq \int_0^{c_1} (z - c_1)^2 f_1(z) dz = \\ \int_0^{c_1} (z - c_1)^2 \frac{3V_1}{c_1^3} z^2 dz &= \frac{V_1 c_1^2}{10}. \end{aligned} \quad (12)$$

Analogously, for the right part consider the parabola $f_2(z) = \frac{h_2}{(c_1 - c)^2} (z - c)^2 = \frac{h_2}{c_2^2} (z - c)^2$ such that $\int_{c_1}^c f_2(y) dy = \int_{c_1}^c g(z) dz = V_2$ (see Figure 6 (b) for an illustration). From $\int_{c_1}^c f_2(y) dy = V_2$ we have that $f_2(z) = \frac{3V_2}{c_2^3} (z - c)^2$. By similar arguments as above in the case of $f_1(z)$, we can show that $g(z)$ and $f_2(z)$ intersect only once, at a point $c'' \in (c_1, c)$. Applying Theorem 5 we have that

$$\begin{aligned} \int_{c_1}^c (z - c_1)^2 g(z) dz &\geq \int_{c_1}^c (z - c_1)^2 f_2(z) dz = \\ \int_{c_1}^c (z - c_1)^2 \frac{3V_2}{c_2^3} (z - c)^2 dz &= \frac{V_2 c_2^2}{10}. \end{aligned} \quad (13)$$

From (12) and (13) we obtain that

$$\begin{aligned} d^2(CH(P), H_{ab}) &= \int_0^{c_1} (z - c_1)^2 g(z) dz + \int_{c_1}^c (z - c_1)^2 g(z) dz \\ &\geq \frac{V_1 c_1^2}{10} + \frac{V_2 c_2^2}{10}. \end{aligned}$$

From the Grünbaum-Hammer-Mityagin theorem, we know that $V_1, V_2 \in [\frac{27}{64}V, \frac{37}{64}V]$. Also, we know that $c_1, c_2 \in [\frac{1}{4}c, \frac{3}{4}c]$. It is not hard to show that, under these constraints, the expression $\frac{V_1 c_1^2}{10} + \frac{V_2 c_2^2}{10}$ achieves its minimum of $\frac{7}{256}Vc^2$ for $V_1 = \frac{27}{64}V, c_1 = \frac{3}{4}c$ or $V_1 = \frac{37}{64}V, c_1 = \frac{1}{4}c$. \square

LEMMA 8. *The variance $d^2(CH(P), H_{ab})$ is bounded from above by $\frac{12729}{71680}Vc^2$.*

PROOF. Without loss of generality, we can assume that $g(z)$ has its maximum in $[c_1, c]$. We split the integral $\int_0^c (z - c_1)^2 g(z) dz$ at c_1 , and prove upper bounds for both parts in the following way: For the left part consider the linear function $f_3(z) = h_3$ such that $\int_0^{c_1} f_3(z) dz = \int_0^{c_1} g(z) dz = V_1$ (see Figure 6 (c) for an illustration). From $\int_0^{c_1} f_3(z) dz = V_1$ we have that $f_3(z) = \frac{V_1}{c_1}$. Since $f_3(z)$ is constant, it intersects $g(z)$ only once, at a point $c' \in (c_1, c)$. By Theorem 5, we have that

$$\begin{aligned} \int_0^{c_1} (z - c_1)^2 g(z) dz &\leq \int_0^{c_1} (z - c_1)^2 f_3(z) dz = \\ \int_0^{c_1} (z - c_1)^2 \frac{V_1}{c_1} dz &= \frac{V_1 c_1^2}{3}. \end{aligned} \quad (14)$$

Now, we are looking for an appropriate function $f_4(z)$ to derive an upper bound on the second part of the integral $\int_{c_1}^c (z - c_1)^2 g(z) dz$. Note that both functions $f_3(z)$ and $f_4(z)$, in general can not be of the type $f(y) = const$, which give us the best upper bound, because it can happen that $f_4(z)$ intersects $g(z)$ twice, and we can not apply Theorem 5. Thus, for the left part we consider the parabola $f_4(z) = \frac{h_4}{c^2} z^2$ such that $\int_{c_1}^c f_4(z) dz = \int_{c_1}^c g(z) dz = V_2$ (see Figure 6 (c) for an illustration). Since $f_4(z)$ and $g(z)$ define the same volume on the interval $[c_1, c]$, they must intersect, and by Proposition 3 we know that if $f_4(z) \neq g(z)$, they can intersect only once, at a point $c' \in (c_1, c)$. Under these conditions, we can apply Theorem 5, and since $f_4(z) = \frac{3V_2}{c^3 - c_1^3} z^2$, we obtain

$$\begin{aligned} \int_{c_1}^c (z - c_1)^2 g(z) dz &\geq \int_{c_1}^c (z - c_1)^2 f_4(z) dz = \\ \int_{c_1}^c (z - c_1)^2 \frac{3V_2}{c^3 - c_1^3} z^2 dz &= \frac{3V_2 c_2^2}{c^2 + c c_1 + c_1^2} \left(\frac{c_2^2}{5} + \frac{c_2 c_1}{2} + \frac{c_1^3}{3} \right). \end{aligned} \quad (15)$$

From (14) and (15) we can conclude that

$$\begin{aligned} d^2(\mathcal{P}, H_{ab}) &= \int_0^{c_1} (z - c)^2 g(z) dz + \int_{c_1}^c (z - c)^2 g(z) dz \\ &\leq \frac{V_1 c_1^2}{3} + \frac{3V_2 c_2^2}{c^2 + c c_1 + c_1^2} \left(\frac{c_2^2}{5} + \frac{c_2 c_1}{2} + \frac{c_1^3}{3} \right). \end{aligned}$$

From the Grünbaum-Hammer-Mityagin theorem, we know that $V_1, V_2 \in [\frac{27}{64}V, \frac{37}{64}V]$. Also, we know that $c_1, c_2 \in [\frac{1}{4}c, \frac{3}{4}c]$. It is not hard to show that, under these constraints, the expression $\frac{V_1 c_1^2}{3} + \frac{3V_2 c_2^2}{c^2 + c c_1 + c_1^2} \left(\frac{c_2^2}{5} + \frac{c_2 c_1}{2} + \frac{c_1^3}{3} \right)$ achieves its minimum of $\frac{12729}{71680}Vc^2$ for $V_1 = \frac{27}{64}V, c_1 = \frac{1}{4}c$. \square

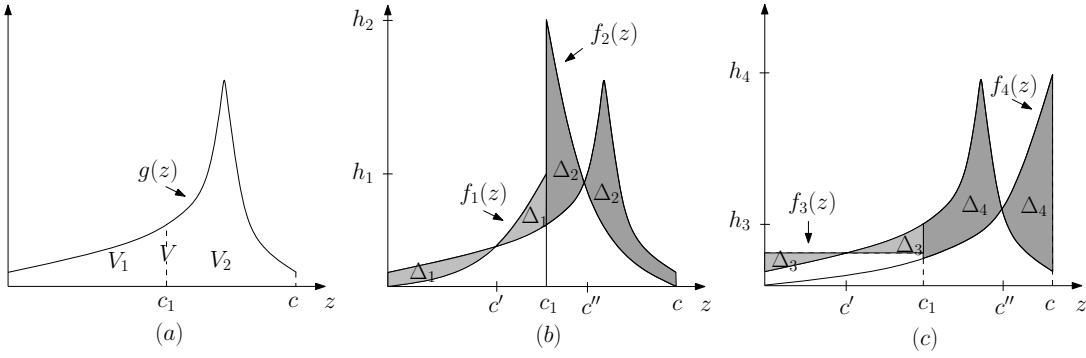


Figure 6: Construction of the lower and upper bounds for $d^2(CH(P), H_{ab})$

So far, we are ready to present a new parametrized bound on $\lambda_{3,3}(P)$, which is good for a large values of η and θ . Additional, crucial relation we exploit in its derivation is the fact given in the following lemma.

LEMMA 9. Let (x_1, x_2, \dots, x_d) and (y_1, y_2, \dots, y_d) be two sets of orthogonal base vectors in \mathbb{R}^d . For any point set $P \in \mathbb{R}^d$ it holds that

$$\sum_{i=1}^d \text{var}(P, x_i) = \sum_{i=1}^d \text{var}(P, y_i).$$

PROOF. We have that

$$\sum_{i=1}^d \text{var}(P, x_i) = \sum_{i=1}^d \frac{1}{n} \sum_{p \in P} d^2(p, H_{x_i}),$$

where H_{x_i} is a hyperplane orthogonal to the vector x_i , passing through the origin of the coordinate system, $d^2(p, H_{x_i})$ denotes the Euclidean distance of p to H_{x_i} , and $n = |P|$. Since $\sum_{i=1}^d d^2(p, H_{x_i})$ is the squared distance of p to the origin of the coordinate system, it can be expressed as the sum of squared distances to the $(d-1)$ -dimensional hyperplanes spanned by any set of orthogonal base vectors. Therefore,

$$\sum_{i=1}^d d^2(p, H_{x_i}) = \sum_{i=1}^d d^2(p, H_{y_i}), \quad \text{and}$$

$$\begin{aligned} \sum_{i=1}^d \text{var}(P, x_i) &= \frac{1}{n} \sum_{p \in P} \sum_{i=1}^d d^2(p, H_{x_i}) = \\ \frac{1}{n} \sum_{p \in P} \sum_{i=1}^d d^2(p, H_{y_i}) &= \sum_{i=1}^d \text{var}(P, y_i). \end{aligned}$$

When P is a continuous point set,

$$\text{var}(P, x_i) = \frac{1}{\text{Vol}(P)} \int_{p \in P} d^2(p, H_{x_i}) ds$$

and the claim can be shown as in the discrete case. \square

LEMMA 10. $\lambda_{3,3}(P) \leq 6.43 \sqrt{1 + \frac{1}{\eta^2} + \frac{1}{\theta^2}}$ for any point set P with aspect ratios $\eta(P) = \eta$ and $\theta(P) = \theta$.

PROOF. Let $x_{pca}, y_{pca}, z_{pca}$ be a set of basis vectors that determine the direction of $BB_{pca(3,3)}(P)$, and let $x_{opt}, y_{opt}, z_{opt}$ be a set of basis vectors that determine the direction of $BB_{opt}(CH(P))$. By Lemma 9, we have that

$$\begin{aligned} \text{var}(CH(P), x_{pca}) + \text{var}(CH(P), y_{pca}) + \text{var}(CH(P), z_{pca}) = \\ \text{var}(CH(P), x_{opt}) + \text{var}(CH(P), y_{opt}) + \text{var}(CH(P), z_{opt}). \end{aligned} \quad (16)$$

By Lemma 1, part i), the variance of $CH(P)$ in the direction x_{pca} is the biggest possible, and therefore

$$\text{var}(CH(P), x_{pca}) \geq \text{var}(CH(P), x_{opt}). \quad (17)$$

Combining (16) and (17) we obtain

$$\begin{aligned} \text{var}(CH(P), y_{pca}) + \text{var}(CH(P), z_{pca}) \leq \\ \text{var}(CH(P), y_{opt}) + \text{var}(CH(P), z_{opt}). \end{aligned} \quad (18)$$

We denote by $H_{a_p b_p}$ the plane orthogonal to z_{pca} , going through the origin of the coordinate system. We assume that the side $a_{pca} b_{pca}$ of $BB_{pca(3,3)}(P)$ lies in $H_{a_p b_p}$. Similarly, we define $H_{a_p c_p}$, $H_{a_o b_o}$ and $H_{a_o c_o}$. We can rewrite (18) as

$$\begin{aligned} d^2(CH(P), H_{a_p b_p}) + d^2(CH(P), H_{a_p c_p}) \leq \\ d^2(CH(P), H_{a_o b_o}) + d^2(CH(P), H_{a_o c_o}). \end{aligned} \quad (19)$$

By Lemma 7, the lower bound for $d^2(CH(P), H_{a_p b_p})$ is $\frac{7}{256} V c_{pca}^2$, and the lower bound for $d^2(CH(P), H_{a_p c_p})$ is $\frac{7}{256} V b_{pca}^2$. By Lemma 8, the upper bound for $d^2(CH(P), H_{a_o b_o})$ is $\frac{12729}{71680} V c_{opt}^2$, and the lower bound for $d^2(CH(P), H_{a_o c_o})$ is $\frac{12729}{71680} V b_{opt}^2$. Plugging these bounds into (19) we obtain

$$\frac{7}{256} V c_{pca}^2 + \frac{7}{256} V b_{pca}^2 \leq \frac{12729}{71680} V c_{opt}^2 + \frac{12729}{71680} V b_{opt}^2. \quad (20)$$

Since $\gamma = \frac{c_{pca}}{c_{opt}}$ and $b_{opt} \geq c_{opt}$, we get from (20) that

$$\beta = \frac{b_{pca}}{b_{opt}} \leq \sqrt{12.99 - \gamma^2}. \quad (21)$$

The expression $\sqrt{12.99 - \gamma^2} \gamma$ ($\geq \beta \gamma$) has its maximum of ≈ 6.437 for $\gamma \approx 2.714$. This together with the bound $\alpha \leq \sqrt{1 + \frac{1}{\eta^2} + \frac{1}{\theta^2}}$ gives

$$\lambda_{3,3}(P) = \alpha \beta \gamma \leq 6.43 \sqrt{1 + \frac{1}{\eta^2} + \frac{1}{\theta^2}}.$$

\square

Lemma 6 gives us a bound on $\lambda_{3,3}(P)$ which is good for small values of η and θ . In contrary, the bound from Lemma 10 behaves worse for small values of η and θ , but better for big values of η and θ . Therefore, we combine both of them to obtain the final upper bound.

THEOREM 8. The PCA bounding box of a point set P in \mathbb{R}^3 computed over $CH(P)$ has a guaranteed approximation factor $\lambda_{3,3} \leq 7.72$.

PROOF. The theorem follows from the combination of the two parametrized bounds from Lemma 6 and Lemma 10:

$$\lambda_{3,3} \leq \sup_{\eta \geq 1, \theta \geq 1} \left\{ \min \left(\eta \theta \left(1 + \frac{1}{\eta^2} + \frac{1}{\theta^2} \right)^{\frac{3}{2}}, 6.43 \sqrt{1 + \frac{1}{\eta^2} + \frac{1}{\theta^2}} \right) \right\}.$$

By numerical verification we obtained that the supremum occurs at ≈ 7.72 . \square

4. FUTURE WORK AND OPEN PROBLEMS

Improving the upper bound for $\lambda_{3,3}$, $\lambda_{2,2}$ and $\lambda_{2,1}$, as well as obtaining an upper bound for $\lambda_{3,2}$ is of interest. A very demanding open problem is to get an approximation factor on the quality of PCA bounding boxes in higher dimensions.

5. REFERENCES

- [1] G. Barequet, B. Chazelle, L. J. Guibas, J. S. B. Mitchell, and A. Tal. Bboxtree: A hierarchical representation for surfaces in 3D. *Computer Graphics Forum*, 15:387–396, 1996.
- [2] G. Barequet and S. Har-Peled. Efficiently approximating the minimum-volume bounding box of a point set in three dimensions. In *10th ACM-SIAM Sympos. Discrete Algorithms*, pages 82–91, 1999.
- [3] N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger. The R*-tree: An efficient and robust access method for points and rectangles. *ACM SIGMOD Int. Conf. on Manag. of Data*, pages 322–331, 1990.
- [4] D. Dimitrov, C. Knauer, K. Kriegel, and G. Rote. Upper and lower bounds on the quality of the PCA bounding boxes. In *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision - WSCG 2007*, pages 185–192, 2007.
- [5] S. Gottschalk, M. C. Lin, and D. Manocha. OBBTree: A hierarchical structure for rapid interference detection. In *SIGGRAPH 1996*, pages 171–180, 1996.
- [6] B. Grünbaum. Partitions of mass-distributions and convex bodies by hyperplanes. *Pacific J. Math.*, 10:1257–1261, 1960.
- [7] I. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 2nd ed., 2002.
- [8] M. Lahanas, T. Kemmerer, N. Milickovic, D. B. K. Karouzakis, and N. Zamboglou. Optimized bounding boxes for three-dimensional treatment planning in brachytherapy. In *Med. Phys.* 27, pages 2333–2342, 2000.
- [9] B. Mityagin. Two inequalities for volumes of convex bodies. *Math. Notes*, 5:61–65, 1968.
- [10] J. O'Rourke. Finding minimal enclosing boxes. In *Int. J. Comp. Info. Sci.* 14, pages 183–199, 1985.
- [11] N. Roussopoulos and D. Leifker. Direct spatial search on pictorial databases using packed R-trees. In *ACM SIGMOD*, pages 17–31, 1985.
- [12] T. Sellis, N. Roussopoulos, and C. Faloutsos. The R+-tree: A dynamic index for multidimensional objects. In *13th VLDB Conference*, pages 507–518, 1987.
- [13] G. Toussaint. Solving geometric problems with the rotating calipers. In *IEEE MELECON*, May 1983.
- [14] D. V. Vranić, D. Saupe, and J. Richter. Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics. In *IEEE 2001 Workshop Multimedia Signal Processing*, pages 293–298, 2001.