


# Support Vector Machines in the Hilbert Geometry

Aditya Acharya ✉ 🏠 

Department of Computer Science, University of Maryland, College Park, MD, USA

Auguste H. Gezalyan ✉ 🏠 

Department of Computer Science, University of Maryland, College Park, MD, USA

Julian Vanecek ✉

Department of Computer Science, University of Maryland, College Park, MD, USA

David M. Mount ✉ 🏠 

Department of Computer Science, University of Maryland, College Park, MD, USA

Sunil Arya ✉ 🏠 

Department of Computer Science and Engineering,  
Hong Kong University of Science and Technology, Hong Kong

---

## Abstract

Support Vector Machines (SVMs) are a class of classification models in machine learning that are based on computing a maximum-margin separator between two sets of points. The SVM problem has been heavily studied for Euclidean geometry and for a number of kernels. In this paper, we consider the linear SVM problem in the Hilbert metric, a non-Euclidean geometry defined over a convex body. We present efficient algorithms for computing the SVM classifier for a set of  $n$  points in the Hilbert metric defined by convex polygons in the plane and convex polytopes in  $d$ -dimensional space. We also consider the problems in the related Funk distance.

**2012 ACM Subject Classification** Theory of computation → Computational geometry

**Keywords and phrases** Support vector machines, Hilbert geometry, linear classification, machine learning, LP-type problems

**Digital Object Identifier** 10.4230/LIPIcs.WADS.2025.3

**Funding** *Sunil Arya*: Research Grants Council of Hong Kong, China project number 16214721.

## 1 Introduction

Support vector machines (SVMs) were introduced by Vapnik and Chervonenkis [34] as a type of classification algorithm [12]. They take two classes of data that are embedded as points in a multidimensional feature space. Once embedded, the algorithm attempts to separate them using a kernel function, which is chosen to maximize the separation distance between the two classes. Throughout, we assume linear kernels, which means that points are separated by a hyperplane.

Two sets of points  $A, B \subset \mathbb{R}^d$  are said to be *linearly separable* if there exists a  $(d - 1)$ -dimensional hyperplane  $h$  such that  $A$  and  $B$  lie on opposite sides of  $h$ , possibly on  $h$  itself (see Figure 1(a)). Define  $\text{Sep}(A, B)$  to be the set of separating hyperplanes. The minimum distance from either set to a separating hyperplane is called its *margin*. Given two linearly separable point sets  $A$  and  $B$  in  $\mathbb{R}^d$ , the (linear) *SVM problem* is that of computing the hyperplane  $h \in \text{Sep}(A, B)$  that maximizes the margin (see Figure 1(b)). The name arises from the fact that, assuming general position, there are  $d + 1$  points that achieve the margin exactly, called the *support vectors*.

SVMs are a classic tool in machine learning [5, 13, 28, 31]. They are highly regarded as a classification method due to their geometric interoperability, robustness against over-fitting, practicality with small and high dimensional datasets, and success in generalizing to unseen



© Aditya Acharya, Auguste H. Gezalyan, Julian Vanecek, David M. Mount, and Sunil Arya;  
licensed under Creative Commons License CC-BY 4.0

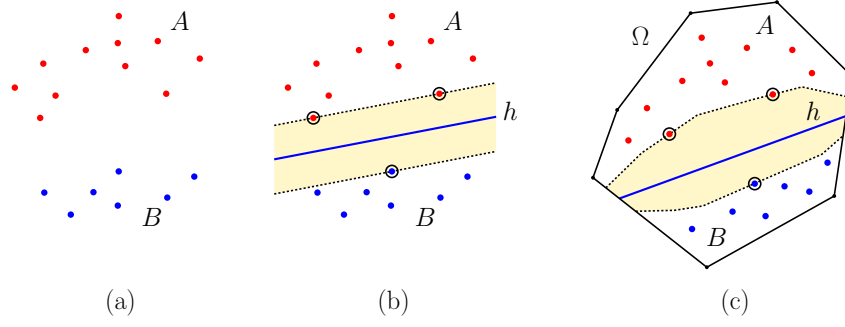
19th International Symposium on Algorithms and Data Structures (WADS 2025).

Editors: Pat Morin and Eunjin Oh; Article No. 3; pp. 3:1–3:20

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



■ **Figure 1** (a) The point sets, (b) Euclidean SVM, (c) and the Hilbert SVM.

data. SVMs are widely used in the fields of vision [23,32], natural language processing [15,22], and computational biology [4,21]. In particular, SVMs have seen a lot of use in image [8,16] and text classification problems [33].

SVMs are well studied in standard Euclidean geometry and other common geometries, such as hyperbolic geometry [10]. They have also been studied relative to a variety of different kernels [20]. In this paper, we consider the SVM problem in a different geometry, called the *Hilbert geometry*. In 1895, David Hilbert introduced the Hilbert metric [19]. It is defined on a convex body  $\Omega$ . The Hilbert metric assigns a distance between two points by considering their relative positions along the chord of  $\Omega$  that passes through them. (Definitions will be provided in Section 2.) The Hilbert metric is widely used in science and engineering. It has many desirable properties, including the facts that it is invariant under projective transformations, straight lines are geodesics, and it generalizes a common model of hyperbolic geometry. An excellent resource on Hilbert geometries is the handbook by Papadopoulos and Troyanov [27]. This is the first work to study SVMs in the Hilbert geometry.<sup>1</sup>

Given the emerging interest in the Hilbert geometry, it is natural to ask whether the classical SVM problem can be efficiently solved in this setting. The input consists of two discrete, linearly separable point sets  $A$  and  $B$  in the interior of a convex body  $\Omega$ , and the objective is to compute a separating hyperplane  $h$  that maximizes the minimum Hilbert distance from every point of  $A \cup B$  to  $h$ . Changing how distances are measured results in very different criteria for the optimum separating hyperplane (see Figure 1(c)). In this paper, we explore these differences and develop an efficient algorithm for the Hilbert SVM problem for convex polygons in the plane. We also generalize our results to the closely related Funk geometry (defined in Section 2.1). The following theorems summarize our main results. For our multidimensional results,  $\Omega$  is a full-dimensional convex polytope of total combinatorial complexity  $m$ . By this, we mean that  $\Omega$  is the convex hull of a finite set of points in general position, and  $m$  is the total number of faces of all dimensions on its boundary.

► **Theorem 1.**

- (i) *The Hilbert SVM problem for a set of  $n$  points in an  $m$ -sided convex polygon in  $\mathbb{R}^2$  is of LP-type and can be solved in  $O(n \log^2 m)$  time.*
- (ii) *For any fixed  $d \geq 3$ , the Hilbert SVM problem for a set of  $n$  points in a convex polytope in  $\mathbb{R}^d$  is of LP-type and can be solved in  $O(nm^3)$  time, where  $m$  is the total combinatorial complexity of  $\Omega$ .*

<sup>1</sup> Note that SVMs have been studied in the context of reproducing kernel Hilbert spaces [37]. Despite the similarity of the names, Hilbert spaces and Hilbert geometries are entirely different.

► **Theorem 2.**

- (i) *The Funk SVM problem for a set of  $n$  points in an  $m$ -sided convex polygon in  $\mathbb{R}^2$  is of LP-type and can be solved in  $O(n \log m)$  time.*
- (ii) *For any fixed  $d \geq 3$ , the Funk SVM problem for a set of  $n$  points in a convex polytope in  $\mathbb{R}^d$  is of LP-type and can be solved in  $O(nm_0)$  time, where  $m_0$  is the number of vertices of  $\Omega$ .*

Note that it is possible to test whether  $A$  and  $B$  are linearly separable through a standard reduction to linear programming. Through an analysis of the geometry of separating hyperplanes in the Funk and Hilbert geometries, we show that in  $\mathbb{R}^d$ , these SVM problems are of LP-type [30] of combinatorial dimension  $d + 1$ . By standard results, in any fixed dimension  $d$ , LP-type problems can be solved with  $O(n)$  violation tests and evaluations of the objective function on point sets of constant size. (See, e.g., Sharir and Welzl [30], Seidel [29], and Clarkson [11], or the deterministic algorithm of Chazelle and Matoušek [9].) Our main technical contribution is a detailed analysis of the geometric structure of the distance between points and lines and hyperplanes in the Funk and Hilbert geometries. We use this analysis to show how to compute the violation tests and objective function evaluations that are required by algorithms for solving LP-type problems.

## 1.1 Applications of the Hilbert Geometry

There are a number of applications of the Hilbert geometry in science and engineering. An example involves discrete probability distributions. A discrete probability distribution  $\Pi = \{p_1, \dots, p_n\}$  can be modeled as a point in the *probability simplex*,  $\{x \in \mathbb{R}^n : x_i \geq 0, \sum_i x_i = 1\}$ . Clearly, this body is equivalent to a simplex in  $\mathbb{R}^{n-1}$ , and hence the  $(n - 1)$ -dimensional simplicial Hilbert distance provides a natural distance measure between two probability distributions. Nielsen and Sun have demonstrated the benefits of using the Hilbert metric over other standard distance functions for probability distributions [25]. They considered clustering tasks in the probability simplex and the correlation ellipsope. They observed that the Hilbert metric performed competitively when compared to other distance models in the probability simplex, such as the Fischer-Hotelling-Rao metric distance and the Kullback-Liebler divergence. They also showed that the Hilbert metric on the probability simplex satisfies the important property of being *information monotone* [25].

Beyond machine learning, the Hilbert metric has found recent uses in the context of convex approximation [2]. This is due to its close relationship with Macbeath regions [1, 35], which, up to scaling factors, are equivalent to Hilbert balls. This has been used to analyze the flag approximability of convex bodies [36] and the construction of efficient convex coverings [3]. Various results from classical computational geometry have been recreated in the Hilbert metric, including algorithms for Voronoi diagrams [6, 18, 18] and Delaunay triangulations [17].

The remainder of the paper is organized as follows. In Section 2, we present preliminary definitions and facts about the Hilbert geometry and SVMs. In Section 3, we present the proof that the Hilbert SVM problem is of LP-type of combinatorial dimension three. In Section 4, we present algorithms for computing the primitives of violation testing and computing the objective function. Finally, in Section 5, we present concluding remarks.

## 2 Mathematical Preliminaries

In this section, we provide a number of definitions and useful facts. Throughout, for our planar results,  $\Omega$  is an  $m$ -sided convex polygon. For our multidimensional results,  $\Omega$  is a full-dimensional convex polytope of total combinatorial complexity  $m$ . By this, we mean that

$\Omega$  is the convex hull of a finite set of points in general position, and  $m$  is the total number of faces of all dimensions on its boundary. Let  $\partial\Omega$  denote its boundary, and let  $\text{int}(\Omega)$  denote its interior. Given two points  $p, q \in \Omega$ , let  $\overline{pq}$  denote the *chord* of  $\Omega$  passing through these points, that is, the intersection of  $\Omega$  with the line through  $p$  and  $q$ . Given a set  $S \subset \mathbb{R}^2$ , let  $\text{conv}(S)$  denote its convex hull. Given two points  $p$  and  $q$ , let  $\overleftrightarrow{pq}$  denote the line that passes through them, that is, their affine closure.

## 2.1 The Funk and Hilbert Distances

In this section, we will introduce the Funk, reverse-Funk, and Hilbert geometries over  $\Omega$ . For any two points  $p, q \in \text{int}(\Omega)$ , if  $p = q$  all the distances are defined to be zero. Otherwise, let  $p'$  and  $q'$  denote the endpoints of the chord  $\overline{pq}$  ordered as  $\langle p', p, q, q' \rangle$  (see Figure 2(a)). These distances are defined in terms of the ratios of distances between pairs of these points.

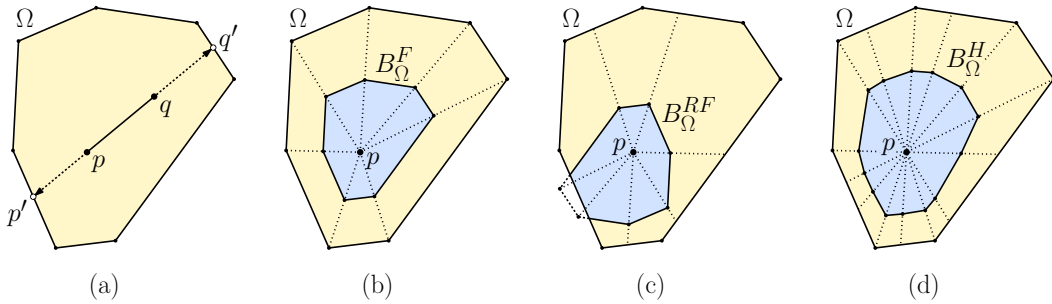
$$\text{Funk: } d_{\Omega}^F(p, q) = \ln \frac{\|p - q'\|}{\|q - q'\|} \quad (1)$$

$$\text{Reverse Funk: } d_{\Omega}^{rF}(p, q) = d_{\Omega}^F(q, p) = \ln \frac{\|q - p'\|}{\|p - p'\|} \quad (2)$$

$$\begin{aligned} \text{Hilbert: } d_{\Omega}^H(p, q) &= \frac{d_{\Omega}^F(p, q) + d_{\Omega}^{rF}(p, q)}{2} \\ &= \frac{1}{2} \ln \left( \frac{\|q - p'\|}{\|p - p'\|} \frac{\|p - q'\|}{\|q - q'\|} \right). \end{aligned} \quad (3)$$

Intuitively,  $d_{\Omega}^F(p, q)$  is a measure of how far  $q$  is from  $p$  relative to the boundary that lies beyond  $q$ , and  $d_{\Omega}^{rF}(p, q)$  is just the reverse of this. The Hilbert distance  $d_{\Omega}^H(p, q)$  symmetrizes these by taking their arithmetic mean.

These distance functions are all nonnegative and satisfy the triangle inequality [27]. The Hilbert distance defines a *metric* over  $\text{int}(\Omega)$ . The Funk and reverse Funk are asymmetric, and so they each define a *weak metric* over  $\text{int}(\Omega)$ . Observe that as either point approaches the boundary of  $\Omega$ , the Hilbert distance approaches  $\infty$ . (This is true for one of the two points in Funk and reverse-Funk.) The Funk and reverse-Funk distances are invariant under invertible affine transformations. The Hilbert distance has the additional property that it is invariant under invertible projective transformations.



■ **Figure 2** (a) Defining distance and balls for (b) Funk, (c) reverse Funk, and (d) Hilbert.

Letting  $d_{\Omega}^*$  denote any of the above and given a point  $p \in \Omega$  and scalar  $r \geq 0$ , the associated metric balls are defined in the standard way.

$$B_{\Omega}^*(p, r) = \{q \in \Omega : d_{\Omega}^*(p, q) \leq r\}.$$

In particular, letting  $\lambda^F(r) = 1 - e^{-r}$  and  $\lambda^{rF}(r) = e^r - 1$ , it is easy to verify that the Funk and reverse-Funk balls are

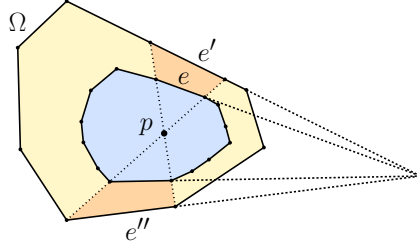
$$B_\Omega^F(p, r) = p + \lambda^F(r)(\Omega - p) \quad (4)$$

$$B_\Omega^{rF}(p, r) = \Omega \cap (p - \lambda^{rF}(r)(\Omega - p)). \quad (5)$$

Thus, the Funk ball is a scaling of  $\Omega$  by a factor of  $\lambda^F(r)$  about  $p$  [26] (see Figure 2(b)), and the reverse-Funk ball is the intersection of  $\Omega$  with a scaling of  $-\Omega$  by a factor of  $\lambda^{rF}(r)$  about  $p$  (see Figure 2(c)). Unlike the Funk ball, this scaling may generally extend outside  $\Omega$ .

The Hilbert ball about a point is a bit more complicated to describe. Let us first consider the planar case. Given a point  $p \in \text{int}(\Omega)$ , define a *spoke* to be the chord  $\overline{pv}$ , for some vertex  $v$  of  $\Omega$ , and define a *half-spoke* to be either of the subsegments of the spoke whose endpoint is  $p$ . The  $m$  vertices of  $\Omega$  define  $m$  spokes, which induce  $2m$  half-spokes joining  $p$  to the boundary of  $\Omega$  (see Figure 2(d)). Nielsen and Shao showed that  $B_\Omega^H(p, r)$  is a convex polygon whose vertices are the  $2m$  points lying on these half-spokes at Hilbert distance  $r$  from  $p$ . Furthermore, the edges of the ball satisfy the following concurrency condition [24].

► **Lemma 3.** *Consider a convex polygon  $\Omega$  in  $\mathbb{R}^2$  and an edge  $e$  on the boundary of some Hilbert ball  $B_\Omega^H(p, r)$ . Let  $e'$  and  $e''$  denote the edges of  $\Omega$  that lie within the double-wedge formed by the two half-spokes that bound  $e$  (see Figure 3). Then the linear extensions of  $e$ ,  $e'$ , and  $e''$  are concurrent.*



■ **Figure 3** Concurrency of Hilbert ball edges.

This lemma generalizes to  $\mathbb{R}^d$ , where  $e$ ,  $e'$  and  $e''$  are facets, but the planar version suffices for our results. Since all distances will be defined with respect to  $\Omega$  henceforth, we will frequently omit explicit references to  $\Omega$  when discussing distances and metric balls.

The concept of spokes can be easily generalized to the case where  $\Omega$  is a convex polytope in  $\mathbb{R}^d$ , and  $p \in \text{int}(\Omega)$ . In particular, for  $0 \leq k \leq \lfloor (d-1)/2 \rfloor$ , consider two faces  $f$  and  $f'$  of  $\Omega$  of respective dimensions  $k$  and  $d-k-1$  such that  $p$  lies in the convex hull of  $f \cup f'$ . It follows from straightforward conditions on dimensionality (see, e.g., [14]) that, assuming general position, for any two such faces, there is at most one line passing through  $p$  that intersects both faces. Define a *spoke* of  $p$  to be the chord of  $\Omega$  contained in any such line, and again a *half-spoke* is the portion of the chord lying on one side of  $p$ . (The planar case arises when  $d = 2$  and  $k = 0$ .)

► **Lemma 4.** *Consider a convex polytope  $\Omega$  in  $\mathbb{R}^d$  of combinatorial complexity  $m$  and a point  $p \in \text{int}(\Omega)$ , for any  $r > 0$ , the Hilbert ball  $B_\Omega^H(p, r)$  is a polytope with  $O(m^2)$  vertices, each of which lies on a half-spoke emanating from  $p$ . When  $d = 2$ , this reduces to  $O(m)$ .*

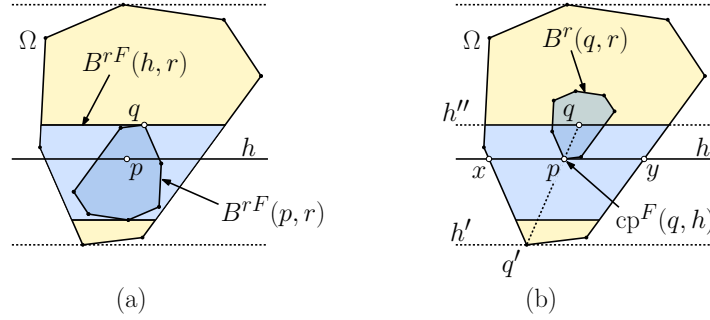
**Proof.** The planar case follows from the analysis of Nielsen and Shao [24]. For the higher-dimensional cases, we appeal to a variational argument. Consider any vertex  $x$  of  $B_\Omega^H(p, r)$ . Suppose to the contrary that the chord  $\overline{px}$  is not a spoke. Consider the points  $y$  and  $y'$

where this chord intersects  $\partial\Omega$ . It follows that there are nonempty affine neighborhoods around  $y$  and  $y'$ , denoted  $N_y$  and  $N_{y'}$ , respectively,  $N_y, N_{y'} \subset \partial\Omega$ , such that lines passing through  $p$  and intersecting  $N_y$  also intersect  $N_{y'}$ . This implies that there is a nonempty affine neighborhood around  $x$  such that  $d_\Omega^H(p, x') = r$ , for all  $x'$  in this neighborhood. Hence,  $x$  is not a vertex of  $B_\Omega^H(p, r)$ . By the observations made above, the number of spokes passing through any point  $p \in \text{int}(\Omega)$  is bounded by the square of the number of faces of  $\Omega$ , that is,  $O(m^2)$ .  $\blacktriangleleft$

## 2.2 On Hilbert and Funk Balls

In this section, we present a number of technical results regarding the geometry and construction of Hilbert and Funk balls for lines and hyperplanes. Due to space limitations, some details are omitted and appear in the full version of the paper.

Let us consider how to compute the reverse-Funk ball of radius  $r$  for a hyperplane  $h$  that intersects a polytope  $\Omega$ . (Later, we will consider how this applies to the planar case, where  $h$  is a line.) Consider any point  $q$  that lies on the boundary of  $B^{rF}(h, r)$  (see Figure 4(a)). This means that the closest point of  $q$  on  $h$  lies at Funk distance  $r$ . Letting  $p$  denote this point, we have  $d^F(q, p) = r$ . Consider the supporting hyperplane  $h'$  of  $\Omega$  that is parallel to  $h$  and on the opposite side from  $q$ . Let  $q'$  be a vertex of  $\Omega$  on  $h'$  (see Figure 4(b)). By convexity, the segment  $\overline{q'q}$  intersects  $h$  at some point  $p$  within  $\Omega$ . Define  $\text{cp}^F(q, h)$  to be this point. By Eq. (4) (noting that  $q$  here plays the role of  $p$  in the equation), we have  $\|p - q\|/\|q' - q\| = \lambda^F(r)$ . It follows that the set of points  $q$  that lie on the boundary of  $B^{rF}(h, r)$  define a hyperplane  $h''$  that is parallel to  $h$ , where the relative distances between these hyperplanes is determined by  $\lambda^F(r)$ . By applying the same logic to the points on the opposite side of  $h$ , it follows that  $B^{rF}(h, r)$  is the intersection of  $\Omega$  with a slab bounded by two hyperplanes that are parallel to  $h$ .



**Figure 4** (a) The reverse-Funk ball around a line and (b) the closest point to  $q$  in the Funk distance.

The times to compute  $\text{cp}^F(q, h)$  and  $d^F(q, h)$  are dominated by the time needed to compute the supporting hyperplanes of  $\Omega$  that are parallel to  $h$ . If  $\Omega$  is an  $m$ -sided convex polygon in  $\mathbb{R}^2$ , this can be done in  $O(\log m)$  time through binary search. If  $\Omega$  is a convex polytope in  $\mathbb{R}^d$  with  $m_0$  vertices, this can be done in  $O(m_0)$  time.

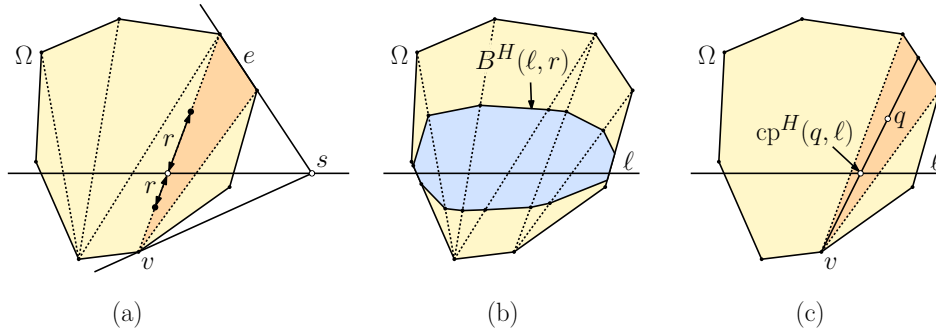
► **Lemma 5.** Consider a convex polytope  $\Omega$  in  $\mathbb{R}^d$  and a  $(d - 1)$ -dimensional hyperplane  $h$  that intersects the interior of  $\Omega$  and  $q \in \text{int}(\Omega)$ . Then

- (i)  $B^{rF}(h, r)$  is the intersection of  $\Omega$  with a slab that is parallel to  $h$ .
- (ii)  $\text{cp}^F(q, h)$  is the point  $p$  on  $h \cap \Omega$  that minimizes the Funk distance  $d^F(q, p)$ .

If  $\Omega$  is an  $m$ -sided convex polygon in  $\mathbb{R}^2$ ,  $\text{cp}^F(q, h)$  and  $d^F(q, h)$  can be computed in  $O(\log m)$  time. If  $\Omega$  is a convex polytope in  $\mathbb{R}^d$  with  $m_0$  vertices, they can be computed in  $O(m_0)$  time.

Next, we consider the Hilbert ball of radius  $r \geq 0$  about a line  $\ell$  that intersects the interior of  $\Omega$ ,  $B^H(\ell, r)$ . The procedure for constructing such a ball was given by Gezalyan and Mount [18, Lemma 6], but we sketch it here for the sake of completeness.

First, extend each of the edges of  $\Omega$  until the extension intersects  $\ell$ . For each such point  $s$ , consider the triangle defined by the convex hull of the edge  $e$  that was extended to form  $s$  and the vertex  $v$  where the supporting line for  $\Omega$  passing through  $s$  hits the opposite side of  $\Omega$  (see Figure 5(a)). These triangles subdivide  $\Omega$ . Each such triangle has two chords that cross  $\ell$ . On each such chord, create two points at Hilbert distance  $r$  from the crossing point with  $\ell$ . These points are in convex position, and by the projective properties of the Hilbert distance, the line segments connecting them through each triangle are part of the Hilbert ball. Together, these points constitute the vertices of the convex hull of the Hilbert ball of the line (see Figure 5(b)).



■ **Figure 5** The Hilbert ball around a line.

Given a point  $q \in \Omega$ , we can compute its closest point on  $\ell \cap \Omega$  as follows. Determine the triangle of the above construction that contains  $q$  (see Figure 5(c)). Consider the chord passing through  $q$  and the appropriate vertex  $v$  of the triangle. Let  $\text{cp}^H(q, \ell)$  denote the intersection point of this chord with  $\ell$ . The chord that joins  $q$  and  $\text{cp}^H(q, \ell)$  is called the *closest-point spoke* for  $q$  with respect to  $\ell$ .

► **Lemma 6.** *Consider a convex polygon  $\Omega$  in  $\mathbb{R}^2$  and a line  $\ell$  that intersects the interior of  $\Omega$  and  $q \in \text{int}(\Omega)$ . Then*

- (i)  $B^H(\ell, r)$  is generated by the above procedure
- (ii)  $\text{cp}^H(q, \ell)$  is the point  $p$  on  $\ell \cap \Omega$  that minimizes the Hilbert distance  $d^H(p, q)$ .

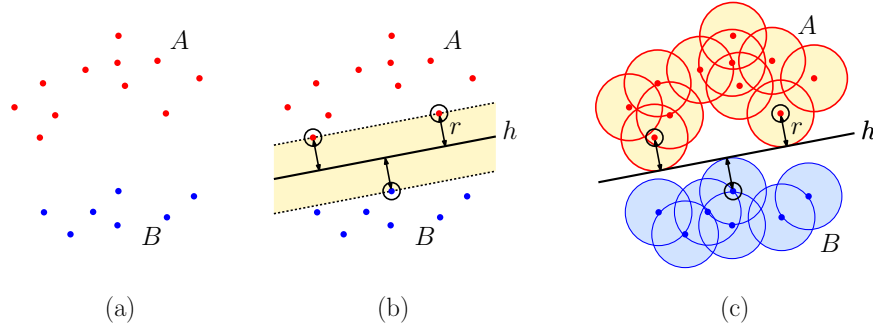
*Assuming  $\Omega$  has  $m$ -sides,  $\text{cp}^H(q, \ell)$  and  $d^H(q, \ell)$  can be computed in  $O(\log^2 m)$  time.*

## 2.3 Support Vector Machines in the Euclidean Geometry

Before discussing SVMs in the Hilbert setting, it is illustrative to explore the differences with the Euclidean case. Given two point sets  $A, B \in \mathbb{R}^d$  (see Figure 6(a)) which are assumed to be linearly separable, the objective is to compute a separating hyperplane  $h$  that maximizes the *margin*, defined to be the minimum distance of either point set to  $h$  (see Figure 6(b)). The points achieving the margin are called the *support vectors*. Local minimality considerations imply that, assuming general position, the number of support vectors does not exceed  $d+1$  [7].

Define a *slab* to be the region bounded between two parallel hyperplanes, the Euclidean SVM problem can be interpreted as the problem of computing the slab of maximum orthogonal width that separates  $A$  and  $B$ . The SVM hyperplane lies midway between the slab's bounding hyperplanes. The margin  $r$  is just half the slab's width (see Figure 6(b)). An alternative and





■ **Figure 6** The Euclidean SVM problem: (a) Point sets  $A$  and  $B$ , (b) the SVM hyperplane and support vectors, (c) separating maximal balls.

equivalent interpretation is to consider the maximum radius  $r$  such that by placing balls of radius  $r$  about the points of  $A$  and  $B$ , the resulting sets of balls are linearly separable (see Figure 6(c)). By replacing Euclidean balls with metric balls, this observation can be readily generalized to any normed space.

### 3 SVMs in the Funk and Hilbert Geometries

In this section, we present our algorithms for solving the SVM problem in the Funk and Hilbert geometries. We consider both the planar and multidimensional cases. Recall that  $\Omega$  is a full-dimensional convex polytope of total combinatorial complexity  $m$ . In both cases, we are given two linearly separable point sets  $A, B \subset \text{int}(\Omega)$  (see Figure 7(a)). The objective is to compute a separating hyperplane  $h$  that maximizes the minimum Hilbert (or Funk) distance from each point of  $A \cup B$  to  $h$ . Throughout, let  $n = |A| + |B|$ . Although we treat both  $n$  and  $m$  as asymptotic quantities, we think of  $m$  more as a large constant, since it defines a fixed geometry.

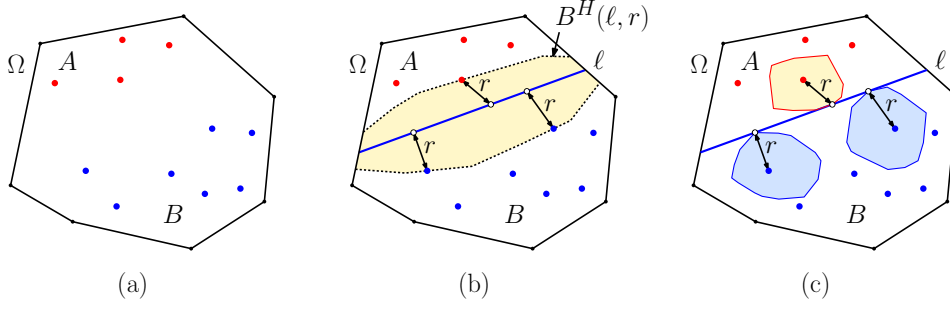
Just as in the Euclidean case (see Section 2.3), there are two equivalent formulations of the problem. In the first, we seek the separating hyperplane  $h$  of maximum radius  $r$  such that  $B^H(h, r)$  contains no points of  $A \cup B$  in its interior (see Figure 7(b)). The other is to determine the maximum radius  $r$  such that the union of balls  $\bigcup_{a \in A} B^H(a, r)$  is linearly separable from  $\bigcup_{b \in B} B^H(b, r)$  (see Figure 7(c)). As in the Euclidean SVM problem, the support vectors are associated with the points of  $A$  and  $B$  whose distance to the SVM separator equals the margin. We augment the concept of support vector to include not only the point but the half-spoke that identifies the vertex of the Hilbert or Funk ball that lies on the separating hyperplane. Note that when dealing with the Funk distance, we use the forward-sense of the distance when computing balls centered at the points, and we use the reverse-Funk when computing the ball centered about the hyperplane.

#### 3.1 Local Optimality

In  $\mathbb{R}^d$ , we refer to the hyperplane  $h$  that achieves the maximum Hilbert margin as the *Hilbert SVM hyperplane* for  $A$  and  $B$ , and we define the *Funk SVM hyperplane* analogously for the Funk distance. The following lemma provides a characterization of this hyperplane.

► **Lemma 7.** *Given a convex polytope  $\Omega$  in  $\mathbb{R}^d$  and two linearly separable point sets  $A, B \subset \Omega$  in general position, the Hilbert (resp., Funk) SVM hyperplane  $h$  satisfies the following condition. Let  $r$  denote the margin achieved by  $h$ . There exist  $d + 1$  pairs  $\{(p_i, s_i)\}_{i=0}^d$ , each*

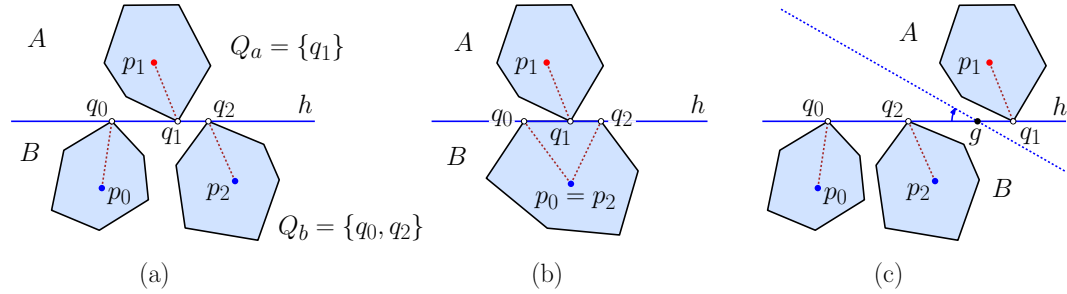




■ **Figure 7** The Hilbert SVM problem.

consisting of a point  $p_i \in A \cup B$  and a half-spoke  $s_i$  of  $p_i$ . For  $0 \leq i \leq d$ , let  $q_i$  denote the unique point at Hilbert (resp., Funk) distance  $r$  from  $p_i$  along  $s_i$ . The points  $Q = \{q_i\}_{i=0}^d$  all lie on  $h$ , and furthermore, if we partition  $Q$  as  $Q_a = Q \cap A$  and  $Q_b = Q \cap B$ , then  $\text{conv}(Q_a) \cap \text{conv}(Q_b) \neq \emptyset$ .

**Proof.** For any  $r \geq 0$ , let  $B^H(A, r) = \bigcup_{a \in A} B^H(a, r)$  denote the set of points of  $\Omega$  whose distance from any point  $A$  is at most  $r$ , and define  $B^H(B, r)$  similarly for  $B$ . As mentioned above, SVM is equivalent to determining the maximum  $r$  such that  $B^H(A, r)$  and  $B^H(B, r)$  are linearly separable. As the value of  $r$  increases towards the optimum margin, the set of hyperplanes that separate these sets of balls becomes progressively smaller, until there is a unique separating hyperplane. Because Hilbert balls are convex polytopes, when this critical event occurs, the separating hyperplane is constrained by  $d + 1$  contact points, which are generally vertices of these balls (see Figure 8(a) and (b)).



■ **Figure 8** Criteria for local optimality for the Hilbert SVM.

By Lemma 4, these vertices lie on the half-spokes of points from  $A$  and  $B$  and each vertex is at distance  $r$  from its respective point. Let  $\{(p_i, s_i)\}_{i=0}^d$  denote these point-spoke combinations, and let  $Q_a$  and  $Q_b$  be as defined in the statement of the lemma. If  $\text{conv}(Q_a) \cap \text{conv}(Q_b) = \emptyset$  (see Figure 8(c)), then there exists a  $(d - 2)$ -dimensional affine subspace  $g \subset h$  that separates them. By rotating  $h$  about  $g$ , we can break the contact with these metric balls, implying that it would be possible to increase the margin, a contradiction. ◀

### 3.2 LP-Type Problems

In this section, we show that our Hilbert and Funk SVM problems possess properties similar to that of low-dimensional linear programming, which leads to an efficient algorithmic solution. In particular, we show that these problems have a property called *LP-type* [30]. This type

of problem requires a finite set  $S$  as input and an objective function  $f$  to be maximized, which induces a total ordering over subsets of  $S$ . Generally, an LP-type problem is defined as follows.

► **Definition 8** (LP-type [30]). *An optimization problem over a set  $S$  and objective  $f$  is of LP-type if:*

- *Monotonicity:* For sets  $S' \subseteq S'' \subseteq S$ ,  $f(S') \geq f(S'') \geq f(S)$ .
- *Locality:* For sets  $S' \subseteq S'' \subseteq S$  and every  $x \in S$ , if  $f(S') = f(S'') = f(S' \cup \{x\})$ , then  $f(S') = f(S'' \cup \{x\})$ .

A *basis* of an LP-type problem is a set  $S' \subseteq S$  with the property that every proper subset of  $S'$  has a larger objective-function value than  $S'$  itself. The *combinatorial dimension* of an LP-type problem is defined to be the maximum cardinality of a basis.

A natural choice for  $S$  would be  $A \cup B$  and a basis would consist of  $d + 1$  points of this set. While this suffices for the simpler Funk distance and the Hilbert distance in the plane, we will need more information for the multidimensional Hilbert SVM problem. In particular, we will use pairs consisting of a point of  $A \cup B$  and a half-spoke for this point. As seen in Lemma 7, these entities locally define the SVM hyperplane. Given  $A$ ,  $B$ , and  $\Omega$ , define  $V = V_\Omega(A, B)$  to be the set of pairs  $(p, s)$ , where  $p \in A \cup B$ , and  $s$  is any half-spoke emanating from  $p$ . Such a pair is called a *point-spoke*. In the multidimensional setting, the role of the set  $S$  in the definition of LP-type will be played by  $V$ .

We define the objective function for the multidimensional Hilbert SVM problem as follows. Recall that  $\text{Sep}(A, B)$  denotes the set of  $(d - 1)$ -dimensional hyperplanes that separate  $A$  and  $B$ . Given a hyperplane  $h \in \text{Sep}(A, B)$  and a point-spoke  $v = (p, s)$ , define  $d_\Omega^H(v, h)$  to be the Hilbert distance from  $p$  to  $h$  measured along the half-spoke  $s$ . If  $s$  does not intersect  $h$ , define the distance to be  $\infty$ . In the planar case, given any set  $S' \subset A \cup B$ , we define  $f(S')$  to be

$$f(S') = \sup_{\ell \in \text{Sep}(A, B)} \min_{p \in S'} d_\Omega^H(p, \ell).$$

In the multidimensional case, given any  $h \in \text{Sep}(A, B)$  and  $V' \subset V$ , define

$$d_\Omega^H(V', h) = \min_{v \in V'} d_\Omega^H(v, h) \quad \text{and} \quad f(V') = \sup_{h \in \text{Sep}(A, B)} d_\Omega^H(V', h).$$

Given  $p \in \text{int}(\Omega)$ , define  $\sigma(p)$  to be the set of half-spokes of  $p \in \text{int}(\Omega)$ . By convexity of Hilbert balls, for any  $p \in A \cup B$  the half-spoke  $s \in \sigma(p)$  that defines the distance  $d_\Omega^H(p, h)$  to any hyperplane  $h$  is the one that minimizes the value of  $d_\Omega^H((p, s), h)$ . Thus, the objective function for our LP-type problem is the following.

$$f(V) = \sup_{h \in \text{Sep}(A, B)} d_\Omega^H(V, h) = \sup_{h \in \text{Sep}(A, B)} \min_{p \in A \cup B} \min_{s \in \sigma(p)} d_\Omega^H((p, s), h).$$

Our next result shows that Hilbert SVM satisfies these conditions.

► **Lemma 9.** *Given a convex polytope  $\Omega$  in  $\mathbb{R}^d$  and a set of linearly separable points  $A, B \subset \Omega$ , the Hilbert SVM problem and the Funk SVM problems are both LP-type problems of combinatorial dimension  $d + 1$ .*

**Proof.** We proved in Lemma 7 that for either distance measure, any optimal solution is determined by a basis consisting of  $d + 1$  point-spoke pairs, which implies that the combinatorial dimension is  $d + 1$ . Monotonicity holds trivially by the fact that adding more point-spokes to a set can never increase the minimum distance from that set to any hyperplane.

To establish locality, suppose that for some sets,  $S' \subseteq S'' \subseteq S$ , we have  $f(S') = f(S'') = f(S' \cup \{x\})$ . We assert that, the maximum-margin separating hyperplane is unique. To see why, suppose to the contrary that  $S'$  and  $S''$  define two distinct separating hyperplanes,  $h'$  and  $h''$ , respectively, both with equal margins. These hyperplanes intersect in a  $(d-2)$ -affine subspace  $g$  (possibly external to  $\Omega$ ). Every strict convex combination  $h$  of these two hyperplanes yields a separating hyperplane that passes through  $g$ . Assuming general position (and particularly that no  $d+1$  point-spokes are equidistant to  $g$  in terms of the Hilbert/Funk distance), it follows that  $h$  does not contact all  $d+1$  basis point-spokes. Since  $d+1$  point-spokes are needed to fix a hyperplane, there exists a perturbation of  $h$  that yields a strictly larger margin, a contradiction.

Therefore, it follows that the support vectors that form this basis are all members of  $S'$ . We may assume that  $x \notin S'$ , since otherwise the conclusion holds trivially. Since  $x \notin S'$ , the metric ball centered at  $x$  whose radius is equal to the margin does not intersect the separating hyperplane. Therefore, adding  $x$  to  $S''$  does not change the margin value, which implies that  $f(S') = f(S'' \cup \{x\})$ , as desired.  $\blacktriangleleft$

## 4 Solving the LP-Type Problems

As mentioned above, in any fixed-dimensional space, an LP-type problem over a ground set of size  $n$  can be solved in  $O(n)$  time assuming free access to two primitives. The first involves evaluating the *objective function* for a given basis set. Second, the *violation test* checks whether a given point causes the objective function to decrease. In this section, we will explore the running times of these primitives.

Given point sets  $A$  and  $B$  of total size  $n$ , recall that the ground set  $S$  for the Funk SVM (in all dimensions) and the planar Hilbert SVM problems consists of  $A \cup B$ . Thus,  $|S| = n$ . For the multidimensional Hilbert SVM, the ground set consists of all point-spoke pairs  $V$ . Assuming that  $\Omega$  has total combinatorial complexity  $m$ ,  $|V| = O(nm)$ . The overall running time to solve the SVM problems will be the product of the size of the ground set and the running time of the two primitives.

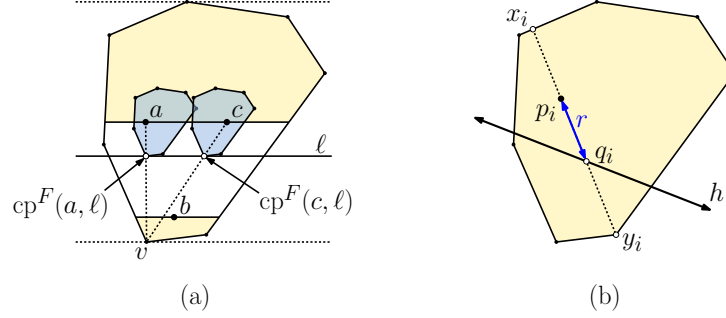
Our algorithms for computing the objective function will return both the SVM hyperplane  $h$  and the margin  $r$  as a witness. This implies that the violation tests are relatively easy. In the case of the Funk distance and the two-dimensional Hilbert distance, we are given a point  $p \in S$  to test. In the Funk case, by Lemma 5, we can compute the distance from  $p$  to  $h$  in  $O(\log m_0)$  time in  $\mathbb{R}^2$  and in  $O(m_0)$  time in the multidimensional case. In the two-dimensional Hilbert case, by Lemma 6, we can compute the distance from  $p$  to  $h$  in  $O(\log^2 m)$  time. In all cases, there is a violation if and only if this distance exceeds  $r$ .

The remaining case is the multidimensional Hilbert geometry. In this case, we are given a point-spoke pair  $(p, s)$  to test. In  $O(m)$  time, we can determine the facets hit by the endpoints of the associated chord. We can then determine the point of intersection of this chord with the hyperplane and compute the distance from  $p$  to the hyperplane along this spoke in constant time. Again, a violation occurs if this distance exceeds  $r$ . All of these tests can be performed within the time bounds needed for the objective function evaluation, so they do not dominate the overall running time.

### 4.1 The Funk Objective Function

We will first discuss the objective function for the Funk SVM, since it is the easier of the two. In the planar case, Lemma 7 implies that a basis consists of three points, say  $\{a, b, c\}$ . From the lemma, we know that two points must come from one set and one from the other, so there

is no loss in generality in assuming that  $a, c \in A$  and  $b \in B$ . Let  $\ell$  be the SVM line we seek. Consider the supporting line of  $\Omega$  that lies on the opposite side of  $\ell$  as  $a$  and  $c$ . Let  $v$  denote the vertex of  $\Omega$  incident to this supporting line. By Lemma 5(ii),  $\text{cp}^F(a, \ell)$  and  $\text{cp}^F(c, \ell)$  lie on the segments  $\overline{av}$  and  $\overline{cv}$ , respectively (see Figure 9(a)). Since  $d^F(a, \ell) = d^F(c, \ell)$ , it follows that  $\ell$  is parallel to  $\overline{ac}$ .



■ **Figure 9** Computing the (a) Funk objective function and (b) Hilbert objective function.

Given the two points  $a$  and  $c$ , we can compute the vertex  $v$  in  $O(\log m)$  time by binary search on the boundary of  $\Omega$ . We can compute the parallel supporting line on the opposite side of  $\Omega$  in  $O(\log n)$  time, which is used to compute  $b$ 's Funk distance. From these two lines, it is straightforward to compute the location of the parallel line  $\ell$  that equalizes the Funk distances among all three points.

By combining this with the observations from Section 4, we can compute both the violation test and the objective function for the planar Funk SVM in time  $O(\log m)$ . Given that the ground set  $A \cup B$  has  $n$  points, the overall running time is  $O(n \log m)$ , and Theorem 2(i) follows.

Next, let us consider the  $d$ -dimensional case, for any fixed  $d \geq 3$ . A basis consists of  $d + 1$  points  $\{p_i\}_{i=0}^d \subseteq A \cup B$ . By Lemma 5(i), the points  $p_i$  lie on two sides of a slab that is parallel to the desired SVM hyperplane. Let  $h$  denote this hyperplane. To compute  $h$ , observe that it can be represented as  $\{x \in \mathbb{R}^d : \langle x, w \rangle = 1\}$ , for some nonzero vector  $w \in \mathbb{R}^d$ . Thus, there exist reals  $\alpha$  and  $\beta$  such that the points  $p_i$  above the slab satisfy  $\langle p_i, w \rangle - \alpha = 0$  and the points below the slab satisfy  $\langle p_i, w \rangle - \beta = 0$ . This yields a system of  $d + 1$  homogeneous linear equations in the  $d + 2$  variables  $\{\alpha, \beta, w_1, \dots, w_d\}$ . We can solve these equations to determine  $w$  up to a nonzero scale factor.

This fixes the orientation of  $h$ , but not its exact location. To compute its location, we first determine the two supporting hyperplanes of  $\Omega$  parallel to  $h$ . This can be done in  $O(m_0)$  time, where  $m_0$  denotes the number of vertices of  $\Omega$ . Given these hyperplanes, we take any basis point of  $A$  and any basis point of  $B$ , and through a straightforward calculation, in  $O(1)$  time we can determine the parallel hyperplane  $h$  that equalizes the Funk distance to each.

In summary, assuming that the dimension  $d$  is fixed, both the violation test and the objective function can be computed in  $O(m_0)$  time. Given that the ground set  $A \cup B$  has  $O(n)$  points, the overall running time is  $O(nm_0)$ . This establishes Theorem 2(ii).

## 4.2 The Multidimensional Hilbert Objective Function

Next, let us consider how to compute the objective function for the  $d$ -dimensional case, for any fixed  $d \geq 3$ . A basis consists of  $d + 1$  point-spoke pairs  $\{p_i, s_i\}_{i=0}^d$ . We claim that we can compute  $h$  in  $O(m)$  time. To see this, observe that  $h$  can be represented as

$\{x \in \mathbb{R}^d : \langle x, w \rangle = 1\}$  for some nonzero vector  $w \in \mathbb{R}^d$ . For a given point-spoke pair  $(p_i, s_i)$ , in  $O(m)$  time we can compute the points  $x_i$  and  $y_i$ , where the extension of the half-spoke intersects  $\partial\Omega$ . Let us assume this is oriented so that  $y_i - x_i$  is in the direction of  $s_i$  (see Figure 9(b)). For a given distance  $r$ , let  $q_i = q_i(r)$  denote the point along this spoke that is at Hilbert distance  $r$  from  $p$ . To simplify the calculation, let  $\hat{r} = \exp(2r)$ , and let us assume that the chord  $\overline{x_i y_i}$  is affinely mapped to the interval  $[0, 1]$  on the real line, which maps  $p_i$  and  $q_i$  to points in this interval (which we will continue to refer to as  $p_i$  and  $q_i$ ). By Eq. (3), for  $0 \leq i \leq d$

$$\hat{r} = \frac{q_i}{p_i} \cdot \frac{1 - p_i}{1 - q_i} \quad \text{or equivalently} \quad \hat{r}(p_i \cdot (1 - q_i)) = q_i \cdot (1 - p_i).$$

Together with the constraints  $\hat{r} \geq 1$  (from the fact that  $r \geq 0$ ) and  $\{0 \leq q_i \leq 1\}_{i=0}^d$  (from the requirement that the  $q_i$ 's lie in the interval  $[0, 1]$ ), this yields a system of  $d + 1$  algebraic equations of degree 2 in the  $d + 1$  unknowns  $\{\hat{r}, w_1, \dots, w_d\}$ . (Note that the  $p_i$  values are fixed.) Since  $d$  is a fixed constant, we can solve this system in  $O(1)$  time. If this system has a feasible solution, then the objective function value is  $r = \frac{1}{2} \ln \hat{r}$ , and the values of the  $q_i$ 's determine the hyperplane by specifying the points where the separating hyperplane crosses each of the half-spokes.

The time to compute the objective function is dominated by the  $O(m)$  time needed to compute the endpoints of each of the half-spoke extensions. By Lemma 4, each point gives rise to  $O(m^2)$  half-spokes, which implies that the ground set  $V$  is of size  $O(nm^2)$ . Thus, the overall running time is  $O(nm^3)$ , and this establishes Theorem 1(ii).

### 4.3 The Hilbert Objective Function in the Plane

Finally, let us consider computing the objective function for the planar Hilbert SVM. From Lemma 7, a basis consists of three support vectors, which must include at least one point from each of  $A$  and  $B$ . As before, if we are presented with a subset that does not meet these conditions, we report a value of  $+\infty$  for the objective function. We will focus on the more interesting case of computing the Hilbert SVM objective function for a basis consisting of three points  $\{a, b, c\}$ . There is no loss in generality in assuming that  $a, c \in A$  and  $b \in B$ . As observed earlier, the value of the objective function is the maximum  $r$  such that the balls  $B(a, r)$  and  $B(c, r)$  are linearly separable from  $B(b, r)$ . The resulting separating line  $\ell$  is equidistant to all three points (see Figure 8(a)). Note that there need not be a separating line that is equidistant to three points (see Figure 8(c)). In this section, we will present an  $O(\log^2 m)$  time algorithm that either finds the desired *equidistant separating line*, or reports that no such line exists. Note that if the line does not exist, the value of the objective function reduces to a two-point case, either  $\{a, b\}$  or  $\{b, c\}$ , whichever achieves the smaller separation distance.

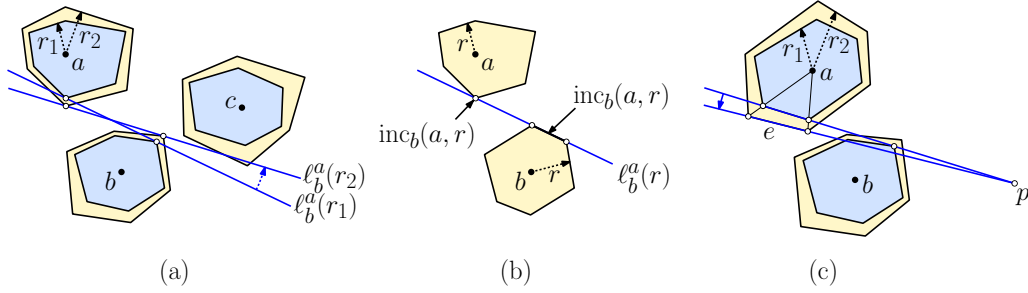
#### 4.3.1 Overview

Let us begin with an overview of the process. For the sake of illustration, we may assume that the points are given in left-to-right order (increasing  $x$ -coordinates) as  $\langle a, b, c \rangle$ , and  $b$  lies vertically below a point on the line segment  $\overline{ac}$ . This implies that there is no vertical separating line, and for any separating line  $a$  and  $c$  will be above it and  $b$  below.

By the observations made before Lemma 6, we know that for any  $q \in \Omega$  and any line  $\ell$ , the distance from  $q$  to  $\ell$  is determined by its closest point  $\text{cp}^H(q, \ell)$ , which lies on the closest-point spoke of  $q$  with respect to  $\ell$  (recall Figure 5(c)). Let  $\ell$  denote the equidistant separating line

(assuming that it exists), and let  $a' = \text{cp}^H(a, \ell)$ , and define  $b'$  and  $c'$  analogously. It suffices to compute  $a'$ ,  $b'$ , and  $c'$ , or, failing this, to report that there is no equidistant separating line for these points.

Our approach will be based on a coordinated binary search. Let us suppose for now that an equidistant line  $\ell$  exists, and let  $r^*$  denote the Hilbert distance of each point from this line. For any radius  $r$ , where  $0 \leq r \leq d^H(a, b)/2$ , let  $\ell_b^a(r)$  denote the unique line that is cross-tangent to the balls  $B^H(a, r)$  and  $B^H(b, r)$ , with  $a$  above and  $b$  below (see Figure 10(a)). Clearly, this line varies continuously as a function of  $r$ . Observe that  $\{c\} = B^H(c, 0)$  lies above  $\ell_b^a(0)$ , and  $c$ 's ball lies above the line until  $r$  reaches the value  $r^*$ , after which this condition fails. This suggests an approach based on a parametric search based on the radius  $r$ .



■ **Figure 10** (a) The cross-tangents  $\ell_b^a(r_1)$  and  $\ell_b^a(r_2)$  for two different radii, (b) definition of  $\text{inc}$ , and (c) the proof of Lemma 10 (uniqueness).

To do this effectively, we need to establish a small set of candidate radii that contains the optimal radius  $r^*$ . We will organize the search around the contact points where the cross-tangent line intersects the boundary of the metric balls. We will show that these boundary points vary monotonically in cyclic order around the center point as the radius increases. Since we know that there is no vertical separating line, we may restrict consideration to the lower half of  $a$ 's and  $c$ 's metric balls (that is, the lower portion of the boundary between the leftmost and rightmost points) and the upper portion of  $b$ 's metric ball.

Recall that for any radius, each vertex of a Hilbert ball lies on one of the half-spokes passing through the center point and each of the  $m$  vertices of  $\Omega$ . Given a point  $q \in \text{int}(\Omega)$ , a vertex of any Hilbert ball centered at  $q$  is uniquely identified by the half-spoke emanating from  $q$  and the ball's radius  $r$ . We refer to this half-spoke as a *virtual vertex*. The *actual vertex* is the point along this half-spoke at distance  $r$  from  $q$ . Similarly, an edge of  $q$ 's metric ball, is uniquely identified by the cone with apex at  $q$  between each pair of consecutive half-spokes. This cone defines a *virtual edge* and, for a given radius  $r$ , the *actual edge* is determined by the line segment between the associated actual vertices.

For any point, there are  $2m$  virtual vertices and  $2m$  virtual edges. Since these entities are associated with elements on the boundary of  $\Omega$ , it is possible to index these entities after  $O(m)$  preprocessing time to split the spokes about the center point into half-spokes and merge them in cyclic order about the center point. Our algorithm involves a cyclic search among the virtual edges and/or the virtual vertices about each metric ball, which is performed by a coordinated binary search. The numerous technical details are presented in Section A.

## 5 Conclusions

In this paper, we presented an efficient algorithm for solving the SVM problem in the Hilbert and Funk geometries for a set of  $n$  points in convex polygons and polytopes of combinatorial complexity  $m$ . Our results rely on the fact that the Hilbert SVM problem has LP-type structure.

Our planar results are quite efficient, but our multidimensional results are polynomial in the total combinatorial complexity of the polytope  $\Omega$ , which generally scales exponentially with the dimension. A natural question is whether the dependency can be reduced, say, to a polynomial in the combinatorial complexity of  $\Omega$ , measured either as the number of facets or the number of vertices. One challenge to this is the issue that the number of vertices on the Hilbert ball grows quadratically in the number of faces of all dimensions of the polytope. Worse yet, the combinatorial structure of the Hilbert ball varies with the location of the center point.

These challenges suggest that the problem may be hard to solve in the case of general polytopes, but there are special cases of the multidimensional Hilbert geometry where the problem is still interesting. For example, Nielsen and Sun have demonstrated the usefulness of the Hilbert metric in applications of machine learning and clustering [25] for the multidimensional simplicial Hilbert geometry. Yet another direction for future work would be to consider kernel shapes other than hyperplanes.

Finally, it would be interesting to know whether our results can be extended to the reverse-Funk distance. Unlike the Funk case, the reverse-Funk ball about a hyperplane is not a slab. It seems reasonable to conjecture that the methods that were used for the more complicated Hilbert geometry could be applied to the case of the reverse-Funk SVM.

---

## References

- 1 A. Abdelkader and D. M. Mount. Economical Delone sets for approximating convex bodies. In *Proc. 16th Scand. Workshop Algorithm Theory*, pages 4:1–4:12, 2018. doi:10.4230/LIPIcs.SWAT.2018.4.
- 2 A. Abdelkader and D. M. Mount. Convex approximation and the Hilbert geometry. In *Proc. SIAM Symp. Simplicity in Algorithms (SOSA24)*, pages 286–298, 2024. doi:10.1137/1.9781611977936.26.
- 3 S. Arya, G. D. da Fonseca, and D. M. Mount. Economical convex coverings and applications. In *Proc. 34th Annu. ACM-SIAM Sympos. Discrete Algorithms*, pages 1834–1861, 2023. doi:10.1137/1.9781611977554.ch70.
- 4 A. Ben-Hur, C. S. Ong, S. Sonnenburg, B. Schölkopf, and G. Rätsch. Support vector machines and kernels for computational biology. *PLoS Comput. Biology*, 4(10):e1000173, 2008. doi:10.1371/journal.pcbi.1000173.
- 5 C. M. Bishop. *Pattern Recogn. and Mach. Learn.* Springer, 2016. doi:10.1007/978-0-387-45528-0.
- 6 M. Bumpus, C. Dai, A. H. Gezalyan, S. Munoz, R. Santhoshkumar, S. Ye, and D. M. Mount. Software and analysis for dynamic Voronoi diagrams in the Hilbert metric, 2023. arXiv:2304.02745.
- 7 C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167, 1998. doi:10.1023/A:1009715923555.
- 8 O. Chapelle, P. Haffner, and V. N. Vapnik. Support vector machines for histogram-based image classification. *IEEE Trans. Neural Netw.*, 10:1055–1064, 1999. doi:10.1109/72.788646.
- 9 B. Chazelle and J. Matoušek. On linear-time deterministic algorithms for optimization problems in fixed dimension. *Journal of Algorithms*, 21:579–597, 1996. doi:10.1006/jagm.1996.0060.



- 10 H. Cho, B. DeMeo, J. Peng, and B. Berger. Large-margin classification in hyperbolic space. In *Proc. 22nd Internat. Conf. Artif. Intell. and Stat.*, pages 1832–1840. PMLR, 2019. doi:10.48550/arXiv.1806.00437.
- 11 K. L. Clarkson. Las Vegas algorithms for linear and integer programming when the dimension is small. *J. Assoc. Comput. Mach.*, 42:488–499, 1995. doi:10.1145/201019.201036.
- 12 C. Cortes and V. Vapnik. Support-vector networks. *Mach. Learn.*, 20:273–297, 1995. doi:10.1007/BF00994018.
- 13 N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.
- 14 S. Das, S. R. Dev, and Sarvottamananda. A worst-case optimal algorithm to compute the Minkowski sum of convex polytopes. *Discrete Appl. Math.*, 350:44–61, 2024. doi:10.1016/j.dam.2024.02.004.
- 15 D. DeCoste and B. Schölkopf. Training invariant support vector machines. *Mach. Learn.*, 46:161–190, 2002. doi:10.1023/A:1012454411458.
- 16 G. M. Foody and A. Mathur. A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans. Geosci. Remote Sensing*, 42:1335–1343, 2004. doi:10.1109/TGRS.2004.827257.
- 17 A. H. Gezalyan, S. Kim, C. Lopez, D. Skora, Z. Stefankovic, and D. M. Mount. Delaunay triangulations in the Hilbert metric. In *Proc. 19th Scand. Workshop Algorithm Theory*, pages 25:1–25:17, 2024. doi:10.4230/LIPIcs.SWAT.2024.25.
- 18 A. H. Gezalyan and D. M. Mount. Voronoi diagrams in the Hilbert metric. In *Proc. 39th Internat. Sympos. Comput. Geom.*, pages 35:1–35:16, 2023. doi:10.4230/LIPIcs.SoCG.2023.35.
- 19 D. Hilbert. Ueber die gerade Linie als kürzeste Verbindung zweier Punkte. *Math. Annalen*, 46:91–96, 1895. doi:10.1007/BF02096204.
- 20 T. Hofmann, B. Schölkopf, and A. J. Smola. Kernel methods in machine learning. *The Annals of Statistics*, 36:1171–1220, 2008. doi:10.1214/009053607000000677.
- 21 S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics & Proteomics*, 15:41–51, 2018. doi:10.21873/cgp.20063.
- 22 T. Joachims. *Learning to Classify Text Using Support Vector Machines*. Springer, 2002. doi:10.1007/978-1-4615-0907-3.
- 23 S. Krig. *Computer Vision Metrics: Survey, Taxonomy, and Analysis*. Springer nature, 2014.
- 24 F. Nielsen and L. Shao. On balls in a Hilbert polygonal geometry (multimedia contribution). In *Proc. 33rd Internat. Sympos. Comput. Geom.*, pages 67:1–67:4, 2017. doi:10.4230/LIPIcs.SoCG.2017.67.
- 25 F. Nielsen and K. Sun. Clustering in Hilbert’s projective geometry: The case studies of the probability simplex and the elliptope of correlation matrices. In Frank Nielsen, editor, *Geometric Structures of Information*, pages 297–331. Springer Internat. Pub., 2019. doi:10.1007/978-3-030-02520-5\_11.
- 26 A. Papadopoulos and M. Troyanov. From Funk to Hilbert geometry. In *Handbook of Hilbert geometry*, volume 22 of *IRMA Lectures in Mathematics and Theoretical Physics*, pages 33–68. European Mathematical Society Publishing House, 2014. doi:10.4171/147-1/2.
- 27 A. Papadopoulos and M. Troyanov. *Handbook of Hilbert geometry*, volume 22 of *IRMA Lectures in Mathematics and Theoretical Physics*. European Mathematical Society Publishing House, 2014. doi:10.4171/147.
- 28 S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition, 2021.
- 29 R. Seidel. Low dimensional linear programming and convex hulls made easy. *Discrete Comput. Geom.*, 6:423–434, 1991. doi:10.1007/BF02574699.

- 30 M. Sharir and E. Welzl. A combinatorial bound for linear programming and related problems. In *Proc. 9th Internat. Sympos. on Theoret. Aspects of Comp. Sci.*, pages 567–579, 1992. doi:10.1007/3-540-55210-3\_213.
- 31 I. Steinwart and A. Christmann. *Support Vector Machines*. Springer, 2008. doi:10.1007/978-0-387-77242-4.
- 32 R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2022. doi:10.1007/978-3-030-34372-9.
- 33 S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *J. Mach. Learn. Res.*, 2:45–66, 2001. doi:10.1162/153244302760185243.
- 34 V. N. Vapnik and A. Y. Chervonenkis. On a class of algorithms of learning pattern recognition. *Automation and Remote Control*, 25(6):937–945, 1964. (In Russian). URL: <http://mi.mathnet.ru/at11678>.
- 35 C. Vernicos. On the Hilbert geometry of convex polytopes. In *Handbook of Hilbert geometry*, volume 22 of *IRMA Lectures in Mathematics and Theoretical Physics*, pages 111–126. European Mathematical Society Publishing House, 2014. doi:10.48550/arXiv.1406.0733.
- 36 Constantin Vernicos and Cormac Walsh. Flag-approximability of convex bodies and volume growth of Hilbert geometries, 2018. arXiv:1809.09471.
- 37 H. Wendland. *Scattered Data Approximation*. Cambridge University Press, 2004. doi:10.1017/CB09780511617539.

## A The Hilbert Objective Function in the Plane

In this section, we present the technical details of the evaluation of the Hilbert SVM objective function in the planar case, which were deferred from the main body of the paper.

### A.1 Incidence Points and Candidate Radii

Define  $\text{inc}_a(b, r)$  to be the incident virtual vertex or virtual edge of  $B^H(a, r)$  with  $\ell_b^a(r)$  (recall Figure 10(b)). (Assuming general position, a line will be incident to a vertex only, but we will be interested in those radii where the line is incident to an edge.) Similarly, define  $\text{inc}_b(a, r)$  to be the incident vertex or edge of  $B^H(b, r)$  with  $\ell_a^b(r)$ . Symmetrically, define  $\ell_b^c(r)$  to be the unique line that is cross-tangent to the balls  $B^H(c, r)$  and  $B^H(b, r)$ , with  $c$  lying above and  $b$  lying below, and define  $\text{inc}_c(b, r)$  and  $\text{inc}_b(c, r)$  analogously. Our approach is to conduct a coordinated binary search among these three sorted lists of incidence entities to determine the particular triple of entities that yield the equidistant separating line.

To limit the search to a discrete set, the candidate radius values will be chosen so that one of the lines  $\ell_b^a(r)$  or  $\ell_b^c(r)$  is incident to an edge of one of the three metric balls. For each virtual edge  $e$  of the metric balls centered at  $a$ , let  $\text{rcand}_a(b, e)$  denote the set of radius values  $r$  such that  $\text{inc}_a(b, r) = e$ . Define  $\text{rcand}_c(b, e)$ ,  $\text{rcand}_b(a, e)$ , and  $\text{rcand}_b(c, e)$  analogously, where  $e$  comes from the appropriate metric ball ( $c$ ,  $b$ , and again  $b$ , respectively). We will show below that each set is either empty or consists of a single element.

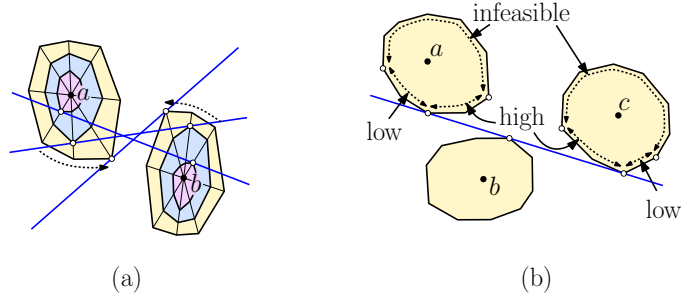
► **Lemma 10.** *Given the point configuration  $\{a, b, c\}$  defined above and any edge  $e$  of  $a$ 's Hilbert ball, if  $\text{rcand}_a(b, e)$  exists, then it is unique. (The same holds for the other point combinations,  $\text{rcand}_c(b, e)$ ,  $\text{rcand}_b(a, e)$ , and  $\text{rcand}_b(c, e)$ , where  $e$  is chosen from the appropriate ball.)*

**Proof.** If  $\text{rcand}_a(b, r)$  exists, then by definition there is a radius  $r$  such that the cross-tangent between  $B_a^H(r)$  and  $B_b^H(r)$  is incident to  $e$  (see Figure 10(c)). Let  $r$  be the smallest such radius and suppose, to the contrary, that there exists a larger radius  $r'$  such that  $r' \in \text{rcand}_a(b, r')$ . Thus, the two cross-tangents  $\ell_b^a(r)$  and  $\ell_b^a(r')$  are both incident on (the virtual edge)  $e$ .

By Lemma 3, every edge of any Hilbert ball centered at a point of  $\Omega$  is concurrent with the point where the linear extensions of two boundary edges of  $\Omega$  meet. The point of concurrency is clearly external to  $\Omega$ . Let  $p$  denote this point. Clearly,  $p$  cannot lie within the portion of either line between the contact points of the lines with either metric ball (which is internal to  $\Omega$ ). As the radius increases, the line rotates about  $p$  in such a way that it moves closer to  $b$ . Since the radius increases, the ball about  $b$  becomes larger, implying that  $\ell_b^a(r')$  intersects  $B_b^H(r)$  (see Figure 10(c)). This contradicts the fact that this line is a cross-tangent between the two balls.

Since this only uses properties of cross-tangents of Hilbert balls, the same proof applies to  $\text{rcand}_c(b, e)$ ,  $\text{rcand}_b(a, e)$ , and  $\text{rcand}_b(c, e)$ . ◀

The second property that we will need to drive a binary search is monotonicity. As the radius increases, the virtual point of incidence moves monotonically around the boundary of the Hilbert ball (see Figure 11(a)). This would be easy if the metric balls were simple scalings of each other (as occurs, for example, with the Funk distance). In the case of Hilbert, our proof exploits the concurrency property of the edges of the metric ball (recall Lemma 3). As in the previous lemma, we will show this for  $\text{inc}_a(b, r)$ , but the argument holds analogously to the others.



■ **Figure 11** (a) The proof of Lemma 11 (monotonicity) and (b) boundary classification.

► **Lemma 11 (Monotonicity).** *Given the point configuration  $\{a, b, c\}$  defined above,  $\text{inc}_a(b, r)$  varies monotonically with  $r$  about the boundary of  $a$ 's metric ball. (The same holds for the other point combinations  $\text{inc}_c(b, r)$ ,  $\text{inc}_b(a, r)$ , and  $\text{inc}_b(c, r)$ , where the motion is counter-clockwise for combinations involving  $a$  and clockwise for combinations involving  $c$ .)*

**Proof.** As the radius  $r$  increases, the cross-tangent line  $\ell_b^a(r)$  changes continuously. Thus, as  $r$  increases, the object of incidence forms a sequence that alternates between the vertices and edges of the metric ball. If there is any instance where monotonicity fails, some edge  $e$  must be visited twice in this sequence. However, this would imply that there are two distinct radius values associated with this edge, which contradicts Lemma 10. ◀

## A.2 Search Procedure

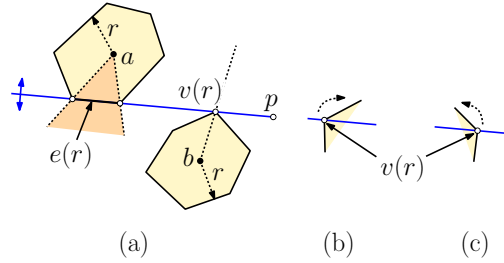
We now have laid the necessary foundations to describe our search. Intuitively, we will label the edges and vertices as being either “low” or “high”, depending on whether a cross-tangent incident to this entity corresponds to a radius value that is smaller than or larger than the optimum value  $r^*$ , respectively. Given an edge  $e$  of  $a$ 's metric ball, if  $r = \text{rcand}_a(b, e)$  exists and the associated cross-tangent  $\ell_b^a(r)$  passes below  $B^H(c, r)$ , we say that the edge is *low*. If the candidate radius exists, but the cross-tangent does not pass below  $B^H(c, r)$ , we say it is *high*. If the candidate radius does not exist, we say that it is *infeasible*.

Each vertex can be categorized similarly. Given a vertex  $v$  of  $a$ 's metric ball, if there exists a radius value  $r$  such that  $\text{inc}_a(b, r) = v$  and  $\ell_b^a(r)$  passes below  $B^H(c, r)$ , we label this vertex as *low*. If there exists such a radius  $r$ , and  $\ell_b^a(r)$  does not pass below  $B^H(c, r)$ , we label it as *high*. Unlike edges, a single vertex may be labeled as both low and high. Finally, if there is no such radius, we label the vertex *infeasible*. By our earlier observation that it suffices to limit consideration to the lower half of  $a$ 's and  $c$ 's metric balls, we can further categorize infeasible points as lying to the left or right of the low-to-high transition point.

As expected, monotonicity implies that if there is an equidistant separating line, the boundary entities of both  $a$ 's and  $c$ 's metric balls have a single transition point from low to high. (Assuming general position, this will take the form of a vertex that is labeled both low and high.) The following lemma summarizes the observations made in this section.

► **Lemma 12.** *Given the point configuration  $\{a, b, c\}$  defined above, if they are not entirely infeasible, the relevant portions of the boundaries of the metric balls for  $a$  and  $c$  consist of a single contiguous low portion followed by a contiguous high portion (possibly sharing a common vertex that is both low and high).*

In light of this lemma, the first task of the search is to find the virtual vertex of  $a$ 's Hilbert ball where the transition from low to high occurs, and then we do the same for  $c$ 's Hilbert ball. Let us consider how this is done for  $a$ , and the algorithm for  $c$  is symmetrical. Consider any (virtual) edge  $e$  of  $a$ 's Hilbert ball. We compute the radius  $r = \text{rcand}_a(b, e)$  by means of a binary search along  $b$ 's metric ball's boundary. Lemma 3 implies that there exists a point  $p$  (external to  $\Omega$ ) about which the virtual edge  $e$  rotates as the radius  $r$  changes. Let  $e(r)$  denote the actual location of this edge for any given radius  $r$  (see Figure 12). Consider any virtual vertex of  $b$ 's Hilbert ball. Define  $v(r)$  to be the actual location of the point along this half-spoke whose distance from  $b$  is  $r$ .



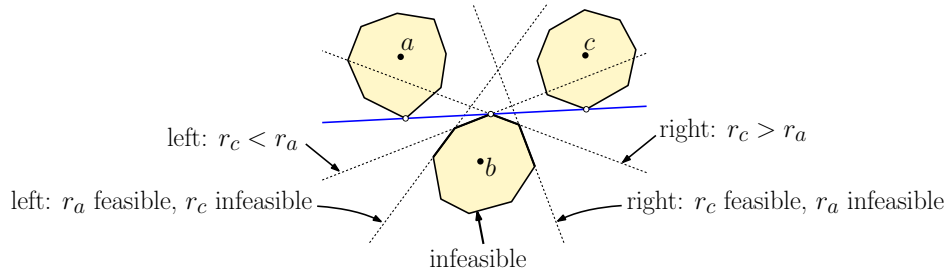
■ **Figure 12** Searching for  $\text{rcand}_a(b, e)$ .

Because  $p$  is external to  $\Omega$ ,  $v(r)$  moves continuously and monotonically along this directed spoke. Therefore, there is at most one value  $r$  where  $v(r)$  lies on the linear extension of  $e(r)$ . If no such  $r$  exists, then this spoke is infeasible for  $e$ . Otherwise, fix  $r$  at the value where  $e(r)$  and  $v(r)$  are collinear. In  $O(1)$  time, we can test whether this common line is a supporting line to  $B^H(b, r)$  that separates the two balls (see Figure 12(a)). If so, then we have successfully found  $r$ . Otherwise, the line intersects  $B^H(b, r)$  in one of two ways, either entering at  $v(r)$  or exiting from  $v(r)$ . Depending on which case applies, monotonicity (Lemma 11) indicates whether we are “low” or “high” with respect to  $b$ 's boundary (see Figure 12(b) and (c)). After  $O(\log m)$  probes to  $b$ 's spokes, either we succeed in finding  $\text{rcand}_a(b, e)$  or we fail. In case of failure, we classify  $e$  as “infeasible”. If it is infeasible, we further classify it as lying to the left or right of the transition point, based on which side of the line  $c$  lies. Since  $b$ 's metric ball is bounded by  $O(m)$  sides and each probe can be performed in  $O(1)$  time, it follows that through binary search, we can compute the values of  $r = \text{rcand}_a(b, e)$ ,  $e(r)$ , and  $v(r)$  in  $O(\log m)$  time.

Once we have established  $r = \text{rcand}_a(b, e)$ , we ask whether the ball  $B^H(c, r)$  lies above or below the line through  $e(r)$  and  $v(r)$ . This reduces to computing the closest point between  $c$  and this line, which, by the observations following Lemma 6, can be accomplished in  $O(\log m)$  time. Based on this, we classify  $e$  on  $a$ 's metric ball as being low or high. Since each probe of an edge of  $a$ 's metric ball takes  $O(\log m)$  time, in  $O(\log^2 m)$  time, we can determine the transition point between low and high edges of  $a$ , whose existence is implied by Lemma 12.

In summary, in  $O(\log^2 m)$  time, we can determine the low-high transition vertex for  $a$ 's metric ball. By repeating the process symmetrically for  $c$  (and the cross-tangent to  $b$ ), we can identify the corresponding vertex for  $c$ 's metric ball.

To complete the process, we need only identify the virtual vertex associated with  $b$ 's metric ball. To do this, we classify the virtual edges of  $b$ 's metric ball as *left*, *right*, or *infeasible*. Let  $e$  be any virtual edge of  $b$ 's metric ball. Apply the same  $O(\log^2 m)$  procedure described above to determine  $r_a = \text{rcand}_b(a, e)$  and  $r_c = \text{rcand}_b(c, e)$ . If both return “infeasible” then  $e$  is labeled as infeasible. If  $r_a$  alone returns a feasible value, we declare  $e$  to be *left* (see Figure 13). If  $r_c$  alone returns a feasible value, we declare  $e$  to be *right*. If both return a feasible value, we classify  $e$  as *left* or *right* depending on whether  $r_a$  is greater than or less than  $r_c$ , respectively. The vertex we desire is the one where the left-to-right transition occurs. The following lemma summarizes the observations made in this section.



■ **Figure 13** Searching  $b$ 's metric ball.

► **Lemma 13.** *Given the point configuration  $\{a, b, c\}$  defined above, if they are not entirely infeasible, the relevant portions of the metric ball boundary for  $b$  consist of a single contiguous left portion followed by a contiguous right portion.*

Based on this lemma, we proceed to compute the vertex that separates the left and right portions in  $O(\log^2 m)$  time, analogous to how we did this for the  $a$  and  $c$  cases. When this is done, we have computed the three virtual vertices for  $a$ ,  $b$ , and  $c$  that define the closest points for the Hilbert SVM line. Thus, we have eliminated the combinatorial element of the problem. All that remains is to solve the remaining numerical problem. The solution is quite technical, but the computation involves only a constant number of entities (six, to be precise), and so the running time is  $O(1)$ . The details are presented in the full version of the paper.

In summary, we have shown that we can compute the violation test in time  $O(\log^2 m)$  and we can evaluate the objective function in time  $O(\log^2 m)$ . Given that both the violation test and the evaluation of the objective function can be performed in  $O(\log^2 m)$  time, this completes the proof of Theorem 1.