

Theorems on the Existence of Separating Surfaces*

Michael E. Houle

School of Computer Science, McConnell Engineering Building, McGill University,
3480 University Street, Montréal, Canada, H3A 2A7

Abstract. Let \mathcal{R} and \mathcal{S} be finite sets in E^d . This paper presents theorems on the existence of strict linear and spherical separators of \mathcal{R} and \mathcal{S} that are similar to the fundamental separation theorem of Kirchberger. Kirchberger's theorem implies that the strict linear separability of finite sets R and G is determined by the separability of all subsets of up to $d + 2$ points of $R \cup G$. This paper shows that under certain conditions, the linear separability of \mathcal{R} and \mathcal{S} is determined by the separability of significantly fewer than all subfamilies of up to $d + 2$ members of $\mathcal{R} \cup \mathcal{S}$. The same treatment is made of Lay's extension of Kirchberger's theorem to separation by hyperspheres.

1. Introduction

Two subsets P and Q of the d -dimensional Euclidean space E^d are said to be (strictly) linearly separable if there exists some hyperplane h such that P is contained in one of the two open half-spaces bounded by h , and Q is contained in the other. In 1903 Kirchberger published a fundamental theorem on the existence of strict linear separators for finite point sets in E^d [5]:

Theorem 1.1 (Kirchberger). *Two finite subsets P and Q of E^d are strictly linearly separable if and only if, for each set T consisting of at most $d + 2$ points of $P \cup Q$, the sets $T \cap P$ and $T \cap Q$ are strictly linearly separable.*

A notion closely related to that of linear separability is spherical separability. Two subsets of E^d are said to be (strictly) spherically separable if there exists some hypersphere s such that the interior of s contains one subset and the exterior of s

* This research was supported by a PGS3 scholarship from NSERC.

contains the other. Lay [6] extended Kirchberger’s theorem to spherical separability in the following manner:

Theorem 1.2 (Lay). *Two finite subsets P and Q of E^d are strictly spherically separable if and only if, for each set T consisting of at most $d + 3$ points of $P \cup Q$ the sets $T \cap P$ and $T \cap Q$ are strictly spherically separable.*

One standard proof of Kirchberger’s theorem, that of Rademacher and Schoenberg [7], employs the well-known theorem due to Helly concerning the existence of points in the common intersection of convex sets [3], [1]. Whereas the original theorem of Helly is somewhat more general, we require only the following restricted formulation:

Theorem 1.3 (Helly). *The members of a finite family \mathcal{K} of convex subsets of E^d have a common intersection point if and only if, for each family \mathcal{F} consisting of at most $d + 1$ members of \mathcal{K} , the members of \mathcal{F} have a common intersection point.*

These theorems are similar in that a “global” property of sets (linear separability, spherical separability, common intersection) is dependent upon the same property considered “locally” over subsets of bounded cardinality, these cardinalities being $d + 2$ for Kirchberger’s theorem, $d + 3$ for Lay’s, and $d + 1$ for Helly’s. It is not difficult to produce examples which demonstrate that the respective cardinalities cannot be decreased using the formulations given above. However, there is still a significant dissimilarity between Helly’s theorem and the others. To illustrate this dissimilarity, let us consider an example. Let $\mathcal{K} = \{K_1, K_2, \dots, K_n\}$ be a family of n convex sets of E^d , $n > d$, defined as follows (see Fig. 1):

1. Sets K_1, K_2, \dots, K_{d+1} are closed half-spaces whose bounding hyperplanes contain the $d + 1$ facets of some d -dimensional simplex in E^d .
2. These half-spaces do not contain the interior of this simplex.
3. The remaining convex sets of \mathcal{K} , if any, are closed balls containing the simplex.

It is easily verified that the members of \mathcal{K} have no point in common, yet with the

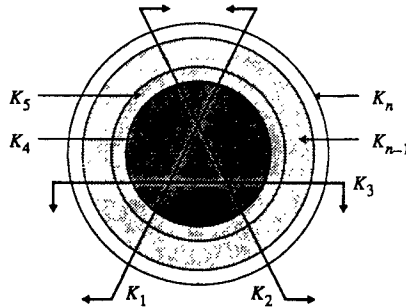


Fig. 1. Convex sets in E^2 with exactly one subfamily of three sets nonintersecting.

exception of the subfamily $\{K_1, K_2, \dots, K_{d+1}\}$, every subfamily consisting of at most $d + 1$ members of \mathcal{K} has a common point of intersection. If we were to test a family of a convex sets for common intersection using Helly's theorem as a guide, we would expect to have to test all $\binom{n}{d+1}$ different subfamilies of cardinality $d + 1$, in the worst case, before being able to make a decision. On the other hand, it is not hard to see that there are no examples of point sets P and Q of E^1 , of combined cardinality $n > 3$, such that P and Q are not linearly separable but only one subset of $P \cup Q$ of cardinality three is not linearly separable. A similar situation exists in the setting of spherical separability. These observations suggest the possibility that Kirchberger's and Lay's theorems are not "optimal," in that fewer than $\binom{n}{d+2}$ subsets of $P \cup Q$ in E^d need be tested for local linear separability in order to ascertain whether P and Q are themselves linearly separable, and fewer than $\binom{n}{d+3}$ subsets need be tested to ascertain whether P and Q are spherically separable. Indeed, this is reflected in the following refinement of Kirchberger's theorem, due to Watson [8]:

Theorem 1.4 (Watson). *Let P and Q be disjoint finite sets of points in E^d , and let x be any point in $P \cup Q$. Sets P and Q are strictly linearly separable if and only if, for each set $T \subseteq P \cup Q$ consisting of at most $d + 2$ points and containing x , the sets $T \cap P$ and $T \cap Q$ are strictly linearly separable.*

The main result of the next section is a generalization of Watson's refinement to finite families of arbitrary subsets of E^d . Two such families \mathcal{R} and \mathcal{S} are said to be (strictly) linearly separable if there exists some hyperplane h such that the member sets of \mathcal{R} are contained in one of the two open half-spaces bounded by h , and the member sets of \mathcal{S} are contained in the other. In an analogous fashion, we may also define the spherical separability of finite families. Section 3 concerns itself with similar treatments of Lay's theorem.

2. Separation Using Hyperplanes

Let $h = \{x \in E^d \mid \pi \cdot x = 1\}$ be a hyperplane avoiding the origin, where $\pi \in E^d, \pi \neq \mathbf{0}$. Of the two open half-spaces delimited by h , we say that the half-space containing the origin, $h^+ = \{x \in E^d \mid \pi \cdot x < 1\}$, be called the *inner* half-space of h . Similarly, the other half-space, $h^- = \{x \in E^d \mid \pi \cdot x > 1\}$, is known as the *outer* half-space of h . Consider the point-hyperplane dual transform \mathcal{D} that maps each point $p \in E^d$ ($p \neq \mathbf{0}$) into the hyperplane $\mathcal{D}(p) = \{x \in E^d \mid p \cdot x = 1\}$, and each origin-avoiding hyperplane $h = \{x \in E^d \mid \pi \cdot x = 1\}$ into the point $\mathcal{D}(h) = \pi$. The following observation is fairly straightforward:

Observation 2.1. Let p be a point E^d other than the origin, and let h be a hyperplane of E^d avoiding the origin. If point p is contained in hyperplane h , then

point $\mathcal{D}(h)$ is contained in hyperplane $\mathcal{D}(p)$. Otherwise, if p is contained in the inner (outer) half-space of h , then $\mathcal{D}(h)$ is contained in the inner (outer) half-space of $\mathcal{D}(p)$.

Let \mathcal{R} and \mathcal{G} be finite families of subsets of the d -dimensional Euclidean space E^d , such that the members of \mathcal{R} and \mathcal{G} are coloured *red* and *green*, respectively. We consider an augmentation \mathcal{D}^* of the dual transform \mathcal{D} that maps red sets $R \in \mathcal{R}$ into a collection of outer half-spaces $\mathcal{D}^*(R)$, and green sets $G \in \mathcal{G}$ into a collection of inner half-spaces $\mathcal{D}^*(G)$. That is, if r is an element of some red set R , then the outer half-space $\mathcal{D}(r)^-$ is a member of $\mathcal{D}^*(R)$; the green case is defined analogously. Since \mathcal{D} is undefined on the origin, we say that a red point at the origin is mapped under \mathcal{D}^* to the empty set \emptyset , and that a green point at the origin is mapped to the entire space E^d . The empty set and the space E^d can be thought of as the outer and inner half-spaces of a hyperplane at infinity, respectively. Finally, if P is a coloured set, we denote the common intersection of the half-spaces of $\mathcal{D}^*(P)$ as $\mathcal{I}(P)$. It should be noted that $\mathcal{I}(P)$ is necessarily convex (possibly empty), as it is the intersection of convex sets. The set $\mathcal{I}(P)$ has an interesting interpretation in light of separation:

Lemma 2.2. *Let P be a green (red) subset of E^d , and let $\mathcal{I}(P)$ be the common intersection of the members of $\mathcal{D}^*(P)$ as defined above. Then point $x \neq \mathbf{0}$ is contained in $\mathcal{I}(P)$ if and only if its dual hyperplane $\mathcal{D}(x)$ has all points of P contained in its inner (outer) half-space.*

Proof. Let h^* be any member of $\mathcal{D}^*(P)$. By definition, h^* is either an inner (outer) half-space of some hyperplane h whose dual point $\mathcal{D}(h)$ is a point of P , or the entire space E^d (empty set \emptyset). If $h^* = E^d$, then the point $\mathcal{D}^*(h^*)$ of P is the origin, and is contained in the inner half-space of every hyperplane that is the dual under \mathcal{D} of points of $\mathcal{I}(P) \setminus \{\mathbf{0}\}$. (If $h^* = \emptyset$, the set $\mathcal{I}(P)$ is empty.) Otherwise, let $x \neq \mathbf{0}$ be a point of E^d contained in h^* . Since x is contained in the inner (outer) half-space of h , by Observation 2.1 we have green (red) point $\mathcal{D}(h)$ contained in the inner (outer) half-space of hyperplane $\mathcal{D}(x)$. Then $h^* \setminus \{\mathbf{0}\}$ is precisely the set of all points of E^d whose dual hyperplanes under \mathcal{D} have inner (outer) half-space containing green (red) point $\mathcal{D}(h)$. Therefore $\mathcal{I}(P) \setminus \{\mathbf{0}\}$ is the set of all points of E^d whose dual hyperplanes under \mathcal{D} have inner (outer) half-spaces containing P . \square

Theorem 2.3. *Let \mathcal{R} and \mathcal{G} be nonempty finite families of subsets of E^d , and let P be any nonempty member of $\mathcal{R} \cup \mathcal{G}$. Then \mathcal{R} and \mathcal{G} are strictly linearly separable if and only if, for each family \mathcal{B} consisting of $d + 1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$, the families $(\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $(\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ are strictly linearly separable.*

Proof. It suffices to prove the nontrivial implication. Without loss of generality, we assume that P is a member of \mathcal{G} and that the members of \mathcal{R} and \mathcal{G} are coloured red and green, respectively. Also without loss of generality, we may translate the sets of \mathcal{R} and \mathcal{G} such that the set P contains the origin. Let \mathcal{B} be a set of $d + 1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$. By assumption, there exists a hyperplane h that separates

the families $\mathcal{B}_R = (\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $\mathcal{B}_G = (\mathcal{B} \cup \{P\}) \cap \mathcal{G}$. Since P contains the origin, hyperplane h must avoid it, and P is contained in the inner half-space of h . Thus all sets of \mathcal{B}_G must be contained in the inner half-space of h , and all sets of \mathcal{B}_R must be contained in the outer half-space. Lemma 2.2 then implies that if B is a member of $\mathcal{B}_R \cup \mathcal{B}_G$, the point $\mathcal{D}(h)$ of E^d is contained in $\mathcal{I}(B)$, which in turn implies that the common intersection of these sets is non-empty. Since every such subset \mathcal{B} of $d + 1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$ has this property, Helly's theorem implies that the common intersection I of all sets of the form $\{\mathcal{I}(Q) | Q \in \mathcal{R} \cup \mathcal{G}\}$ is nonempty.

It should be noted that I does not contain the origin: otherwise, since no outer half-space may contain the origin, the set \mathcal{R} would be empty, violating the assumption. Let $x \neq \mathbf{0}$ be a point of I . Since x is contained in $\mathcal{I}(Q)$ for each $Q \in \mathcal{R} \cup \mathcal{G}$, Lemma 2.2 again implies that each member R of \mathcal{R} is contained in the outer half-space of hyperplane $\mathcal{D}(x)$, and each member G of \mathcal{G} is contained in the inner half-space. Therefore the sets \mathcal{R} and \mathcal{G} are strictly linearly separable as required. \square

The open half-spaces of a linear separator for families \mathcal{R} and \mathcal{G} may be labeled according to the family contained by each. In the context of Theorem 2.3, this labeling involves a degree of freedom that is eliminated by the choice of some distinguished set P of $\mathcal{R} \cup \mathcal{G}$. In this sense, P acts as a “focus” or a “reference” for the local tests of linear separability. The next theorem shows that we may refer to a distinguished *direction* instead of a distinguished set.

For simplicity of exposition, we assume that the distinguished direction is that of the positive x_d -axis, and refer to it as the *vertical* direction. A hyperplane h that does not contain a translate of the x_d -axis is said to be *nonvertical*. The open half-spaces of h can be described analytically as $h^+ = \{x \in E^d | x_d > \sum_{i=1}^{d-1} \pi_i x_i + \pi_d\}$ and $h^- = \{x \in E^d | x_d < \sum_{i=1}^{d-1} \pi_i x_i + \pi_d\}$. The half-spaces h^+ and h^- are called the *upper* and *lower* half-spaces of h , respectively. The points of h^+ are said to be *above* h , and the points of h^- are said to be *below*.

Theorem 2.4. *Let \mathcal{R} and \mathcal{G} be nonempty finite families of subsets of E^d . Then \mathcal{R} and \mathcal{G} are strictly separable by a nonvertical hyperplane with \mathcal{R} above the hyperplane and \mathcal{G} below if and only if, for each family \mathcal{B} consisting of $d + 1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$, the families $\mathcal{B} \cap \mathcal{G}$ are strictly separable by a nonvertical hyperplane with $\mathcal{B} \cap \mathcal{R}$ above and $\mathcal{B} \cap \mathcal{G}$ below.*

Proof. Let \mathcal{B} be a family of $d + 1$ or fewer members of $\mathcal{R} \cup \mathcal{G}$ as defined above, and let h_b be a nonvertical hyperplane such that $\mathcal{B} \cap \mathcal{R}$ is above h_b and $\mathcal{B} \cap \mathcal{G}$ is below. Also, let P be the intersection of all upper half-spaces of hyperplanes h_b over all finitely many choices of subfamily \mathcal{B} of $\mathcal{R} \cup \mathcal{G}$. Note that P cannot be empty. Then the families $(\mathcal{B} \cup \{P\}) \cap (\mathcal{R} \cup \{P\})$ and $(\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ are linearly separable. Therefore the families $\mathcal{R} \cup \{P\}$ and \mathcal{G} are strictly linearly separable by Theorem 2.3. But every vertical hyperplane intersects P , so the separator must be nonvertical. Finally, P being above the linear separator implies the result. \square

3. Separation Using Hyperspheres

In the proof of this theorem on spherical separability, Lay transforms an instance of a spherical separability problem in E^d into a linear separability problem in E^{d+1} , by means of a stereographic projection. In this new setting, Lay applies Kirchner's theorem directly to obtain his result. In this section we adapt Lay's proof in proving existence theorems for spherical separators similar to the linear separation theorems of the previous section.

Let h be a hyperplane in E^d , and let σ be a hypersphere tangent to h at point p . Let p' be the point of σ antipodal to p . The *stereographic projection* τ of point $x \in h$ onto σ (based at p') is defined as being the intersection of the line containing x and p' with $\sigma \setminus \{p'\}$. This establishes a bijective correspondence between points of h and points of $\sigma \setminus \{p'\}$. Before presenting the theorems of this section, we present (without proof) some basic properties of stereographic projections. For additional information on stereographic projections and geometric transformations in general, the reader is referred to [2] and [4].

Lemma 3.1. *Let h be a hyperplane in E^{d+1} and let σ be a d -dimensional hypersphere of unit radius tangent to h at point p . Let τ be the stereographic projection of h onto σ based at the point p' antipodal to p in σ . Let s be a $(d-1)$ -dimensional sphere contained in h , and let s^+ and s^- be its interior and exterior relative to h , respectively. Then:*

1. *The projection $\tau(s)$ of s onto σ is the intersection of σ and some hyperplane h_s .*
2. *The projections $\tau(s^+)$ and $\tau(s^-)$ are each contained in different open half-spaces defined by h_s .*

See Fig. 2 for an illustration of these relationships. A flat f of dimension $d-1$ contained in h may be viewed as a degenerate $(d-1)$ -dimensional sphere centered at infinity with infinite radius. Obviously, the stereographic projection of f is contained in the intersection of σ and a hyperplane passing through both p' and f .

Theorem 3.2. *Let \mathcal{R} and \mathcal{G} be nonempty finite families of subsets of E^d , and let P be any nonempty member of $\mathcal{R} \cup \mathcal{G}$. Then \mathcal{R} and \mathcal{G} are strictly separable by a (possibly degenerate) hypersphere if and only if, for each family \mathcal{B} consisting of $d+2$ or fewer*

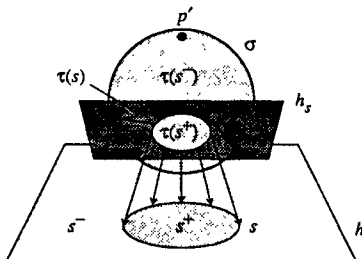


Fig. 2. The stereographic projection of a $(d-1)$ -dimensional sphere.

members of $\mathcal{R} \cup \mathcal{G}$, the families $(\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $(\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ are strictly separable by a (possibly degenerate) hypersphere.

Proof. Let E^d be embedded in some hyperplane h of E^{d+1} , and let σ be a d -dimensional unit sphere tangent to h at some arbitrary point p . Let τ be the stereographic projection of h onto σ based at the point of p' antipodal to p . Let \mathcal{B} be a family consisting of $d + 2$ or fewer members of $\mathcal{R} \cup \mathcal{G}$ such that $\mathcal{B}_R = (\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $\mathcal{B}_G = (\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ are spherically separable in h by some $(d - 1)$ -dimensional sphere s . If h_s is a hyperplane containing $\tau(s)$, then Lemma 3.1 implies that the families $\tau(\mathcal{B}_R)$ and $\tau(\mathcal{B}_G)$ are strictly linearly separable by h_s . Therefore, by Theorem 2.3, the families $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$ are linearly separable.

Let h'_s be a linear separator of $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$ such that h'_s intersect σ in some $(d - 1)$ -dimensional sphere s' . Since $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$ are both nonempty, such a separator must exist. Then the (possibly degenerate) $(d - 1)$ -dimensional sphere $\tau^{-1}(s')$ strictly separates \mathcal{R} and \mathcal{G} . \square

Figure 3 gives an example of two families of sets in E^2 where every subfamily of five members is strictly spherically separable, but the only separator for the entire collection is degenerate. It should be noted that the closure of the triangles of Fig. 3 intersect the separator h , but the triangles themselves do not. In the formulation of the previous theorem, if we restrict the members of \mathcal{R} and \mathcal{G} to be compact (closed and bounded) sets, we can guarantee the nondegeneracy of the separating hyperspheres:

Theorem 3.3. *Let \mathcal{R} and \mathcal{G} be nonempty finite families of compact subsets of E^d , and let P be any nonempty member of $\mathcal{R} \cup \mathcal{G}$. Then \mathcal{R} and \mathcal{G} are strictly spherically separable if and only if, for each family \mathcal{B} consisting of $d + 2$ or fewer members of $\mathcal{R} \cup \mathcal{G}$, the families $(\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $(\mathcal{B} \cup \{P\}) \cap \mathcal{G}$ are strictly spherically separable.*

Proof. As in Theorem 3.2, we embed E^d into a hyperplane h of E^{d+1} and apply a stereographic projection, arriving at a hyperplane h'_s that separates $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$. If h'_s contains the previously defined point p' , then, due to the compactness of $\tau(\mathcal{R})$ and $\tau(\mathcal{G})$, we may perturb h'_s into some new separator h'' that avoids p' . If s'' is the intersection of h'' and σ , then the $(d - 1)$ -dimensional sphere $\tau^{-1}(s'')$ is a nondegenerate strict separator for \mathcal{R} and \mathcal{G} . \square

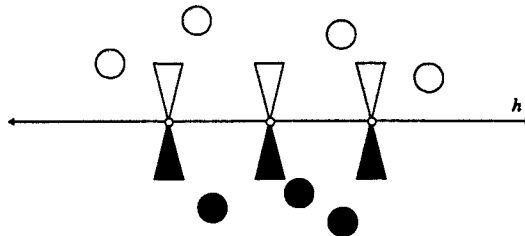


Fig. 3. The only spherical separator is degenerate.

In Theorem 2.4 the need for a distinguished set P of $\mathcal{R} \cup \mathcal{G}$ for linear separability was obviated by the introduction of a distinguished direction. In the setting of spherical separability, this distinction of direction becomes more natural. Let s^+ and s^- be the open interior and exterior of hypersphere s , respectively. We say that the points of s^+ are *inside* s , and that the points of s^- are *outside* s . We now state a theorem of spherical separability analogous to Theorem 2.4.

Theorem 3.4. *Let \mathcal{R} and \mathcal{G} be nonempty finite families of subsets of E^d . Then \mathcal{R} and \mathcal{G} are strictly separable by a hypersphere with \mathcal{R} inside the hypersphere and \mathcal{G} outside if and only if, for each family \mathcal{B} consisting of $d + 2$ or fewer members of $\mathcal{R} \cup \mathcal{G}$, the families $\mathcal{B} \cap \mathcal{R}$ and $\mathcal{B} \cap \mathcal{G}$ are strictly separable by a hypersphere with $\mathcal{B} \cap \mathcal{R}$ inside and $\mathcal{B} \cap \mathcal{G}$ outside.*

Proof. Let \mathcal{B} be a family of $d + 2$ or fewer members of $\mathcal{R} \cup \mathcal{G}$ as defined above, and let s_b be a hypersphere whose interior contains $\mathcal{B} \cap \mathcal{R}$ and whose exterior contains $\mathcal{B} \cap \mathcal{G}$. Also, let s_p be a hypersphere whose interior contains s_b for all finitely many choices of subfamily \mathcal{B} of $\mathcal{R} \cup \mathcal{G}$. If P is the exterior of s_p , then the families $(\mathcal{B} \cup \{P\}) \cap \mathcal{R}$ and $(\mathcal{B} \cup \{P\}) \cap (\mathcal{G} \cup \{P\})$ are spherically separable. Therefore the families \mathcal{R} and $\mathcal{G} \cup \{P\}$ are (possibly degenerately) strictly spherically separable by Theorem 3.2. But every hyperplane intersects P , so the separator cannot be degenerate. Finally, P being outside the spherical separator implies the result. \square

Acknowledgments

The author would like to thank Mike Parker, Thomas Shermer, and the referees for their constructive criticism during the preparation of this paper, Rephael Wenger for helpful discussions relating to its topic, and Godfried Toussaint for providing the starting point.

References

1. Danzer, L., Grünbaum, B., and Klee, V., Helly's theorem and its relatives, in *Convexity*, pp. 101–181, Proceedings of Symposia in Pure Mathematics, Vol. 7, American Mathematical Society, Providence, RI, 1963.
2. Eves, H. W., *A Survey of Geometry*, rev. edn., Allyn and Bacon, Boston, 1972.
3. Helly, E., Über Mengen konvexer Körper mit gemeinschaftlichen Punkten, *Jahresber. Deutsch. Math.-Verein.* **32** (1923) 175–176.
4. Iaglom, I. M., *Geometric Transformations*, tr. A. Shields, Random House, New York, 1962.
5. Kirchberger, P., Über Tchebychefsche Annäherungsmethoden, *Math. Ann.* **57** (1903) 509–540.
6. Lay, S. R., On separation by spherical surfaces, *Amer. Math. Monthly* **78** (1971) 1112–1113.
7. Rademacher, H., and Schoenberg, I. J., Helly's theorem on convex domains and Tchebycheff's approximation problem, *Canad. J. Math.* **2** (1950) 245–256.
8. Watson, D., A refinement of theorems of Kirchberger and Carathéodory, *J. Austral. Math. Soc.* **15** (1973) 190–192.

Received July 27, 1988, and in revised form December 5, 1989.