HELLY'S THEOREMS ON CONVEX DOMAINS AND TCHEBYCHEFF'S APPROXIMATION PROBLEM

HANS RADEMACHER AND I. J. SCHOENBERG

1. Introduction. Professor Dresden called to our attention the following theorem:[1]

If $S_1, S_2, \ldots, S_m$ are m line segments parallel to the y-axis, all of equal lengths, whose projections on the x-axis are equally spaced, and if we assume that a straight line can be made to intersect every set of three among these segments, then there exists a straight line intersecting all the segments.

This theorem was conjectured by M. Dresher; a first proof, unpublished, was communicated to us by T. E. Harris. Wide generalizations are possible. Dr. Harris noticed that we can dispense with the equidistance of the lines carrying our segments. We shall see in a moment that the equality of the lengths of the segments is likewise a superfluous assumption. A further generalization, also due to Dr. Harris, is as follows: The intersecting straight lines can be replaced by general parabolic curves

$$y = a_0x^n + a_1x^{n-1} + \ldots + a_n \qquad (n \leqslant m - 2); \tag{1}$$

again, if each set of $n + 2$ among our segments can be cut by such a parabola, then all may be simultaneously intersected by one such curve.

In this note we wish to point out the close connection of this problem, and of the more general problem of best approximation in the sense of Tchebycheff, with two remarkable theorems on convex domains, due to E. Helly, which may be stated as follows;[2]

THEOREM 1 (Helly). *If $C_1, C_2, \ldots, C_m$ is a finite collection of convex sets, which need not be closed or bounded, in the n-dimensional Euclidean space E_n $(m \geqslant n + 1)$, such that every $n + 1$ among the sets have a common point, then all m sets have a common point.*

THEOREM 2 (Helly). *Let $\{D\}$ be an infinite collection of closed and convex sets D, which need not be bounded, in E_n, such that every $n + 1$ among the sets have a common point. Then all the sets D have a common point, provided there exists a finite subcollection $D', D'', \ldots, D^{(k)}$, $(k \geqslant 1)$, of elements of $\{D\}$, such that their intersection $\Delta = D'D'', \ldots, D^{(k)}$ is non-void and bounded.*

Let us first see how very directly the Dresher-Harris theorem may be derived from Helly's Theorem 1. Let

$$S_\nu:\quad x = x_\nu,\quad b_\nu \leqslant y \leqslant c_\nu$$
$$(\nu = 1, \ldots, m;\quad x_1 < x_2 < \ldots < x_m,\quad m \geqslant 3),$$

Received March 22, 1949.

[1]See [8], p. 4, where the theorem is stated without proof.

[2]See [4], [6], and [7].

https://doi.org/10.4153/CJM-1950-022-8 Published online by Cambridge University Press

be the m segments of the theorem. Consider the totality of lines $y = a_0x + a_1$ intersecting the νth segment S_ν; the requirement of intersection is expressed by the inequalities

$$b_\nu \leqslant a_0x_\nu + a_1 \leqslant c_\nu. \tag{2}$$

In the plane E_2 of the variables (a_0,a_1), the double inequality (2) defines a parallel strip of slope $- x_\nu$. This strip, which we denote by C_ν, is certainly a convex set in E_2. Let us now consider the collection of m convex sets C_1, C_2, . . . , C_m, corresponding to the segments S_1, S_2, . . . , S_m. From the assumption of the Dresher-Harris theorem we know that every three among these sets have a common point. Since all assumptions af Helly's Theorem 1 are satisfied in E_2, we may conclude that all m sets C_ν have a common point, hence all m segments S_ν are intersected by a line. This proof clearly extends to the case of the parabolic curves (1) by applying Helly's Theorem 1 in the space E_{n+1} of the variables $(a_0, a_1, \ldots, a_n)$.

As a second example of the versatility of Helly's ideas we shall again use Theorem 1 to give a new proof of the following separation theorem.[3]

THEOREM OF PAUL KIRCHBERGER. *Let $S = \{P\}$ and $S' = \{P'\}$ be two finite sets of points in E_n. We shall say that a hyperplane π separates strictly S from S', if all points of S are on one side of π, while all points of S' are on the other side, with none of the points lying on π. Such a strictly separating plane π exists if and only if the following condition is satisfied: For every set T of $n + 2$ points chosen arbitrarily from S and S', there should exist a hyperplane π_T which separates strictly the S-points of T from the S'-points of T.*

The necessity of the condition is obvious; to prove its sufficiency let us assume that it is satisfied and prove the existence of a strictly separating plane. We introduce in E_n a coordinate system $(x_1, \ldots, x_n)$. In the space E_{n+1} of the variables $(a_1, a_2, \ldots, a_{n+1})$, and corresponding to each point $P = (x_1, \ldots, x_n)$ of S, we define an open half-space H_P by the inequality

$$H_P:\quad a_1x_1 + a_2x_2 + \ldots + a_nx_n + a_{n+1} > 0. \tag{3}$$

Likewise, corresponding to each point $P' = (x'_1, x'_2, \ldots, x'_n)$ of S', we define in E_{n+1} an open half-space $H_{P'}$ by the inequality

$$H_{P'}:\quad a_1x'_1 + a_2x'_2 + \ldots + a_nx'_n + a_{n+1} < 0. \tag{4}$$

[3]See [5], where a proof of this theorem requires nearly 24 pages. The theorems of Kirchberger and Dresher-Harris are not unrelated. The following new generalization of the Dresher-Harris theorem indicates the connection: *Let S be a finite set of points $P_i = (x_i,y_i)$ in the plane and let S' be a second set of points $P'_j = (x'_j,y'_j)$. We say that a line $y + a_0x + a_1$ separates the sets S and S', if $y_i \geqq a_0x_i + a_1$ for all points of S, and $y'_j \leqq a_0 x'_j + a_1$ for all points of S'. There exists a line $y = a_0x + a_1$ separating S from S' if and only if the following condition is satisfied: For every set T of three points chosen from $S + S'$ there should exist a line separating the S-points of T from the S'-points of T.* We obtain the Dresher-Harris theorem as a special case of this theorem if we take S to be the set of upper endpoints of the segments S_ν, while S' is the set of their lower endpoints. A proof of this generalization by means of Theorem 1 is obvious and so is its extension to parabolic curves (1).

In terms of the finite collection $\{H_P\} + \{H_{P'}\}$ of open half-spaces of E_{n+1}, Kirchberger's condition means that every $n + 2$ among these convex half-spaces have a common point. By Helly's Theorem 1 we conclude that there is a point $(a_1, \ldots, a_{n+1})$, with $\sum_1^n |a_\nu| > 0$, common to all of these half-spaces.[4] The corresponding plane $a_1x_1 + \ldots + a_nx_n + a_{n+1} = 0$ separates strictly S from S', and the proof is concluded. That Kirchberger's theorem becomes false if the number $n + 2$ is replaced by $n + 1$ is seen by the example of the set S being the set of $n + 1$ vertices of a simplex, while S' has only one element namely the centroid of the simplex. These two sets S, S' cannot be separated, though the points of S and S' occurring in any $(n + 1)$-tuple can be separated. We also wish to remark that the theorem becomes false if the sets S, S' are allowed to be infinite. Indeed, if in E_2 we take S to be the exponential curve $y = \exp x$ while S' is the x-axis, then clearly every $n + 2 = 4$ points of $S + S'$ can be strictly separated by a line, but not the sets S, S'.

Concerning Kirchberger's theorem the following remark is of interest. Let us replace the "strict separation" of the theorem by "separation" in the weaker sense that points of S or S' are also allowed to lie *on* the separating plane π. We may then state the following proposition;

Kirchberger's theorem in E_n remains true if in its statement "strict separation" is replaced by "separation" in the above wider sense, provided we replace in the theorem's condition the number $n + 2$ by $2n + 2$. Also no number smaller than $2n + 2$ will do. Moreover the sets S and S' may now also be infinite.

In order to prove this new result let us define in E_{n+1}, as we did above, the collection of *closed* half-spaces

(5) $\quad \overline{H}_P$: $a_1x_1 + \ldots + a_nx_n + a_{n+1} \geqslant 0$, for $P = (x_1, \ldots, x_n) \in S$,

(6) $\quad \overline{H}_{P'}$: $a_1x'_1 + \ldots + a_nx'_n + a_{n+1} \leqslant 0$, for $P' = (x'_1, \ldots, x'_n) \in S'$.

We wish to prove the existence of a point $(a_1, \ldots, a_{n+1})$, with $\sum_1^n |a_\nu| > 0$, which is common to *all* of these half-spaces. Helly's theorem does not help us here any more, but we can apply the following remarkable theorem of L. L. Dines and N. H. McCoy:[5]

A finite or infinite collection of closed half-spaces in E_{n+1}, each half-space having the origin 0 on its boundary, do have a common point different from 0, if every set of $2n + 2$ among our half-spaces have a common point different from 0.

This theorem assures the existence of a point $(a_1, \ldots, a_{n+1}) \neq (0, \ldots, 0)$ such that the inequalities (5) and (6) hold for all $P \in S$, and all $P' \in S'$,

[4]Actually we know only that $\sum_1^{n+1} |a_\nu| > 0$. However, all the points of sufficiently small spherical neighbourhood of the point $(a_0, \ldots, a_{n+1})$ likewise satisfy all conditions and among them we can certainly find one for which $\sum_1^n |a_\nu| > 0$.

[5]See [3], pp. 61-63; see also [2], pp. 962-963, where there are also references to a paper by C. V. Robinson.

respectively. This would mean that $a_1x_1 + \ldots + a_nx_n + a_{n+1} = 0$ is a separating hyperplane as soon as we know that $\sum_1^n |a_\nu| > 0$. This last point, however, is clear, for $a_1 = \ldots = a_n = 0$ and (5) and (6) would imply $a_{n+1} \geqslant 0$, $a_{n+1} \leqslant 0$, hence $a_{n+1} = 0$, which is impossible.

The following example shows that the number $2n + 2$ of the new version of Kirchberger's theorem may not be replaced by $2n + 1$: Let S consist of the $n + 1$ vertices $P_1, \ldots, P_{n+1}$ of simplex σ, and let S' consist of the same $n + 1$ points $P'_1, \ldots, P'_{n+1}$, with $P'_\nu = P_\nu$. Choosing $2n + 1$ points of $S + S'$ amounts to leaving out P_ν, or perhaps P'_ν. The remaining $2n + 1$ points are clearly separated by the $(n - 1)$-dimensional face of the simplex σ which is opposite to the vertex P_ν. Hence the conditions of the theorem are verified for every set of $2n + 1$ points, while there is no hyperplane π separating S from S'.

The connection of Helly's theorems with the idea of Tchebycheff approximation, i.e. the consideration of the minimum of a maximum, suggested to us a new proof[6] which we claim to be the first proof of Helly's theorems to be entirely geometric, in the sense that every single one of its steps has an intuitive geometric meaning. This proof is given in the first part of the paper. The second and last part is devoted to an application of Helly's theorems to Tchebycheff's approximation problem.

A New Proof of Helly's Theorems

2. On proximity points of convex domains. We shall see that the main point in proving Helly's theorems is to prove Theorem 1 for the special case when the convex sets $C_1, \ldots, C_m$ are also *closed* and *bounded.* A closed and bounded convex set in E_n will be referred to as a *convex domain.* Let $D_1, D_2, \ldots, D_m$ be such convex domains. If $Q \in E_n$, we denote by $d(Q,D_\nu)$ the distance from the point Q to the domain D_ν. The point function

$$f(Q) = \max_\nu d(Q,D_\nu)$$

is evidently non-negative and continuous throughout E_n. Since $f(Q) \to \infty$ as $Q \to \infty$, the function $f(Q)$ assumes somewhere its absolute minimum value.

Definition. *An absolute minimum point P, of $f(Q)$, will be called proximity point of our domains $D_1, \ldots, D_m$. It has the property*

$$\max_\nu d(P,D_\nu) = \min_Q \max_\nu d(Q,D_\nu).$$

[6]Three earlier proofs have come to our attention: By J. Radon [7], E. Helly [4], and D. König [6]. Radon's proof, which is the shortest, is analytic. The proofs by Helly and König, essentially equivalent to each other, are geometric. However, all three proofs use the method of mathematical induction, a fact which seems to obscure the intuitive background of the results. Our proof uses the *metric* of E_n and is therefore related to the ideas of Menger and Blumenthal (see [2]).

The minimal value $f(P) = \min f(Q)$ *will be called the proximity of these domains and denoted by the symbol* Prox $(D_1, \ldots, D_m)$.

Evidently we have Prox $(D_1, \ldots, D_m) = 0$ if and only if our m domains have a point in common and if this is the case, any point of their intersection is a proximity point.[7]

3. A characteristic property of proximity points.[8] The following theorem expresses a fundamental property of proximity points.

THEOREM 3. *Let* $D_1, \ldots, D_m$, $(m \geqslant 2)$, *be convex domains in* E_n *having no common point. Let* P *be a proximity point of these domains, the proximity* $\rho = \text{Prox}\,(D_1, \ldots, D_m)$ *being necessarily positive. Let* $P_\nu \in D_\nu$ *be such that* $PP_\nu = d(P,D_\nu)$, *hence we have* $\rho = \max PP_\nu$. *Then there are* s *among the* m *normals* PP_ν *from* P *to our domains,* $PP_1, PP_2, \ldots, PP_s$, *say, such that*

(i) $2 \leqslant s \leqslant n + 1$.

(ii) $PP_1 = PP_2 = \ldots = PP_s = \rho$.

(iii) *The points* $P_1, P_2, \ldots, P_s$ *are the vertices of a* $(s - 1)$*-simplex* σ, *which simplex contains the point* P *in its* $(s - 1)$*-dimensional interior.*

(iv) *The corresponding* s *domains* $D_1, D_2, \ldots, D_s$ *have no common point.*

The last conclusion is for us the important one. In fact we shall use this theorem only in the following abbreviated form:

COROLLARY. *If* m *convex domains, of* E_n, *have no common point, then some* s *among these domains have no common point, where* $2 \leqslant s \leqslant n + 1$.

Proof of Theorem 3. Suppose that

$$PP_1 = PP_2 = \ldots = PP_h = \rho, \quad PP_{h+1} < \rho, \ldots, PP_m < \rho \qquad (h \leqslant m).$$

Clearly $h \geqslant 2$; for if $h = 1$, then P could not be a proximity point. Indeed, then $\max d(P,D_\nu)$ could be diminished below its present value ρ by moving P slightly along PP_1 towards P_1.

Consider now the convex hull of the points $P_1, \ldots, P_h$, which we denote by

$$K = K(P_1, \ldots, P_h).$$

We claim that $P \in K$, for otherwise let PP' be the shortest distance from P to K; we could then, again as before, diminish *all* distances $PP_\nu = d(P,D_\nu)$, $(\nu = 1, \ldots, h)$, by moving P slightly along PP' towards P'. Hence indeed $P \in K(P_1, \ldots, P_h)$.

We shall now use the following known result:[9] *If* P *is a point of the convex*

[7]As illustrations of the notion of a proximity point we mention the following two propositions of elementary geometry: *Let* A, B, C, *be the vertices of an acute-angled triangle in the plane. The proximity point of the three points* A,B,C, *is the circumcenter of the triangle. The proximity point of the three segments* BC, CA, AB, *is the incenter.*

[8]The properties (i), (ii), and (iii) of the point P, as described in Theorem 3 are indeed *characteristic* for a proximity point, a fact which we mention without proof because we do not use it.

[9]See [1], Satz IX on p. 607. This theorem is easily derived from a well known result of Caratheodory to the effect that every point of $K(P_1, \ldots, P_h)$ is a centroid with positive masses of at most $n + 1$ points among the P_ν.

hull $K(P_1, \ldots, P_h)$, then either P coincides with one of the points P_ν, or else we can find a simplex σ, of dimension $s-1$ ranging from 1 to at most n, having as vertices only points from among the points P_ν, and such that P is in the $(s-1)$-dimensional interior of σ. Returning to our proximity point P, we remark that P cannot possibly coincide with any of the points P_ν, since $PP_\nu = \rho > 0$, $(\nu = 1, \ldots, h)$. Therefore the above result assures us of the existence of a simplex of vertices $P_1, P_2, \ldots, P_s$, say, satisfying the conditions (i), (ii), and (iii), of Theorem 3.

There remains to prove the fourth and last statement of the theorem to the effect that $D_1D_2, \ldots, D_s = \phi$. We consider the s unit vectors

$$\vec{a}_i = \overrightarrow{PP_i}/PP_i \qquad (i = 1, \ldots, s),$$

and the s half-spaces H_i defined by

$$H_i\colon\ \overrightarrow{PQ}\,.\,\vec{a}_i \geqslant \rho \qquad (i = 1, \ldots, s). \tag{7}$$

Since $D_i \subset H_i$, it is sufficient to show that

$$H_1H_2\ldots H_s = \phi. \tag{8}$$

Suppose (8) were false and let $Q \in H_i$, $(i = 1, \ldots, s)$; then all inequalities (7) hold. However, since P is in the interior of σ, we have a vector relation of the form

$$\sum_1^s \kappa_i\vec{a}_i = \vec{0}, \text{ with all } \kappa_i > 0. \tag{9}$$

But then, on multiplying (9) scalarly by $\overrightarrow{PQ}$, in view of (7), we obtain

$$0 = \sum \kappa_i(\overrightarrow{PQ}\,.\,\vec{a}_i) \geqslant \sum \kappa_i\rho = \rho\sum\kappa_i,$$

which clearly contradicts the positivity of ρ and κ_i. This completes our proof.

4. A proof of Helly's Theorem 1 concerning a finite collection of convex sets. We distinguish two cases.

First case: We assume that the m convex sets C_ν of the theorem are also *closed* and *bounded*, an assumption which we emphasize by writing $C_\nu = D_\nu$, $(\nu = 1, \ldots, m)$. This case is now immediately disposed of, for if we assume to the contrary that our m convex domains D_ν have no common point, then, by the Corollary of Theorem 3, some s among them $(2 \leqslant s \leqslant n+1)$, have no point in common, a fact which contradicts the assumption of Theorem 1 to the effect that every $n+1$ domains have a common point.

Second case. We assume that the C_ν are convex sets which need not be closed or bounded. By assumption every combination $C_{i_0}, C_{i_1}, \ldots, C_{i_n}$ of $n+1$ distinct sets have a common point. Let such a point be $A_{i_0, i_1, \ldots, i_n}$ and let it be regarded as a symmetric function of its $n+1$ distinct subscripts. Corresponding to each C_i we now define the convex domain

$$D_i = K(A_{i, j_1, \ldots, j_n})$$

which is defined as the convex hull of the $\binom{m-1}{n}$ points $A_{i,\ j_1, \ldots, j_n}$ where $j_1, \ldots, j_n$ runs over all combinations of n among the $m - 1$ numbers $1, \ldots, i-1, i+1, \ldots, m$. Since $A_{i, j_1, \ldots, j_n} \in C_i$, we have

$$D_i \subset C_i, \tag{10}$$

because C_i is convex. Every set of $n + 1$ among these domains, $D_{i_0}, D_{i_1}, \ldots, D_{i_n}$, say, have a point in common, namely the point $A_{i_0, i_1, \ldots, i_n}$. By the *first case* already established we conclude that all D_i have a common point. In view of (10) we now obtain the desired conclusion to the effect that the sets C_i have a point in common.

5. A proof of Helly's Theorem 2 concerning an infinite collection of closed convex sets. Let $\{D\}$ be the given infinite collection of closed convex sets. By Theorem 1 we know that the elements of every finite subcollection of $\{D\}$ have a common point. Consider the new collection $\{D^* = \Delta D\}$, where Δ is the non-void and bounded set defined in the statement of Theorem 2, while D ranges over the given collection $\{D\}$. The elements of $\{D^*\}$ have the following properties;

(i) They are closed, bounded and *non-void* convex sets.

(ii) The elements of every finite set of D^*'s have a common point. The desired conclusion to the effect that all the D^*'s, and therefore also all the D's, have a common point now follows from the following general Theorem:

THEOREM OF F. RIESZ:[10] *If a collection $\{A\}$ of bounded and closed sets in E_n has the property that the elements of every finite subcollection have a common point, then all A's have a common point.*

AN APPLICATION OF HELLY'S THEOREM TO TCHEBYCHEFF'S APPROXIMATION PROBLEM[11]

6. Approximations to discontinuous functions. We first derive somewhat differently a classical result concerning the following finite problem: Let there be given $n + 1$ points

$$(x_\nu, y_\nu) \qquad (\nu = 0, 1, \ldots, n; \quad x_0 < x_1 < \ldots < x_n); \tag{11}$$

we wish to determine the polynomial

$$P(x) = a_0 x^{n-1} + a_1 x^{n-2} + \ldots + a_{n-1} \tag{12}$$

which minimizes the expression

$$\max_\nu |y_\nu - P(x_\nu)|. \tag{13}$$

We need the following lemma: *If the real variables $(u_0, u_1, \ldots, u_n)$ are connected by the linear relation with real constant coefficients*

[10]For the first published proof of Riesz's theorem see [6], p. 210; it is an almost immediate consequence of the Heine-Borel theorem.

[11]An excellent reference to Tchebycheff's approximation problem is [9], Chapter VI.

(14) $$b_0 u_0 + b_1 u_1 + \ldots + b_n u_n = c \qquad (b_0 b_1 \ldots b_n \neq 0),$$
then the expression $\max\limits_{\nu} |u_\nu|$ *has the minimal value*

(15) $$\rho = \frac{|c|}{|b_0| + |b_1| + \ldots + |b_n|},$$

which is reached for just one set of values $u_\nu = u^*_\nu$ *given by*

(16) $$u^*_\nu = \rho \operatorname{sgn}(cb_\nu) \qquad (\nu = 0, \ldots, n).$$

We lose no generality in assuming that $c > 0$, for if $c = 0$ the result is trivial and if $c < 0$ we may multiply both sides of (14) by -1. In view of (15), the relations (16) indeed define a solution of (14); (16) also imply that $|u^*_\nu| = \rho$, hence $\rho = \max |u^*_\nu|$. Let now (u_ν) be an arbitrary set satisfying the two relations

$$\sum b_\nu u_\nu = c \quad \text{and} \quad \max_\nu |u_\nu| \leqslant \rho.$$

This set (u_ν) must be of the form

$$u_\nu = \epsilon_\nu u^*_\nu = \epsilon_\nu \rho \operatorname{sgn}(b_\nu), \text{ where } -1 \leqslant \epsilon_\nu \leqslant 1, \qquad (\nu = 0, \ldots, 1).$$

Now $c = \sum b_\nu u_\nu = \sum b_\nu \epsilon_\nu \rho \operatorname{sgn}(b_\nu) = \sum \epsilon_\nu \rho |b_\nu|$, or $\sum e_\nu \rho |b_\nu| = c$. In view of (15) or $\sum \rho |b_\nu| = c$, the last relation implies that $\epsilon_\nu = +1$ for all ν, and therefore $u_\nu = u^*_\nu$. This completes the proof of our lemma.

Returning to the problem of minimizing (13), let $u_\nu = y_\nu - P(x_\nu)$ be the discrepancies between the points (11) and the polynomial (12). These discrepancies are not independent variables, for they are obviously connected by the single linear relation

$$\begin{vmatrix} u_0 - y_0 & 1 & x_0 \ldots x_0^{n-1} \\ u_1 - y_1 & 1 & x_1 \ldots x_1^{n-1} \\ \ldots & & \ldots \\ u_n - y_n & 1 & x_n \ldots x_n^{n-1} \end{vmatrix} = 0,$$

or

$$\begin{vmatrix} u_0 & 1 \ldots x_0^{n-1} \\ u_1 & 1 \ldots x_1^{n-1} \\ & \ldots \\ u_n & 1 \ldots x_n^{n-1} \end{vmatrix} = \begin{vmatrix} y_0 & 1 \ldots x_0^{n-1} \\ y_1 & 1 \ldots x_1^{n-1} \\ & \ldots \\ y_n & 1 \ldots x_n^{n-1} \end{vmatrix}.$$

Since the coefficients of (u_ν) on the left-hand side of this linear relation alternate in sign, we obtain by (15) for the minimal value of (13) the explicit expression

(17) $$\rho = \text{abs. val.} \begin{vmatrix} y_0 & 1 & x_0 \ldots x_0^{n-1} \\ y_1 & 1 & x_1 \ldots x_1^{n-1} \\ & \ldots & \\ y_n & 1 & x_n \ldots x_n^{n-1} \end{vmatrix} \div \begin{vmatrix} 1 & 1 & x_0 \ldots x_0^{n-1} \\ -1 & 1 & x_1 \ldots x_1^{n-1} \\ & \ldots & \\ (-1)^n & 1 & x_n \ldots x_n^{n-1} \end{vmatrix}.$$

By (16) we also know that this minimal value ρ is reached for *just one* polynomial $P(x)$ for which the discrepancies $u_\nu = y_\nu - P(x_\nu)$ are all equal in absolute value to ρ and alternate in sign.

Concerning the analytical problem of best approximation of functions we wish to prove the following

THEOREM 4. *Let $f(x)$ be a real function defined in $\alpha \leqslant x \leqslant \beta$ about which we only assume that it is bounded. Given n ($n \geqslant 1$), there exists a real polynomial $P^*(x)$, of degree not exceeding $n - 1$, which minimizes the expression*

$$(18) \qquad \sup_{\alpha \leqq x \leqq \beta} |f(x) - P(x)|,$$

giving it its minimal value

$$\rho = \sup_{\alpha \leqq x \leqq \beta} |f(x) - P^*(x)|.$$

For this minimal value ρ we have the relation

$$(19) \qquad \rho = \sup_{(x_\nu)} \rho(x_0, x_1, \ldots, x_n) \quad \textit{for} \quad \alpha \leqslant x_0 < x_1 < \ldots < x_m \leqslant \beta,$$

where

$$(20) \quad \rho(x_0,x_1,\ldots,x_n) = \text{abs. val.} \begin{vmatrix} f(x_0) & 1 & x_0 & \ldots & x_0^{n-1} \\ f(x_1) & 1 & x_1 & \ldots & x_1^{n-1} \\ & & \ldots & & \\ f(x_n) & 1 & x_n & \ldots & x_n^{n-1} \end{vmatrix} \div \begin{vmatrix} 1 & 1 & x_0 & \ldots & x_0^{n-1} \\ -1 & 1 & x_1 & \ldots & x_1^{n-1} \\ & & \ldots & & \\ (-1)^n & 1 & x_n & \ldots & x_n^{n-1} \end{vmatrix}$$

In words, (19) means that the best approximation ρ of our function is the supremum of its best approximation $\rho(x_0, \ldots, x_n)$ over sets of $n + 1$ distinct points of the range $[\alpha,\beta]$.

Proof. Since $f(x)$ is bounded, so are the best approximations (20). Let

$$(21) \qquad \rho_0 = \sup_{(x_\nu)} \rho(x_0,x_1, \ldots, x_n) \quad \text{for} \quad \alpha \leqslant x_0 < x_1 < \ldots < x_n \leqslant \beta.$$

In the space E_n of the variables $(a_0,a_1, \ldots, a_{n-1})$ and corresponding to each value of x in the range $[\alpha,\beta]$, we consider the parallel layer of space D_x defined by

$$(22) \qquad D_x\colon |f(x) - a_0x^{n-1} - a_1x^{n-2} - \ldots - a_{n-1}| \leqslant \rho_0.$$

We claim that the collection $\{D_x\}$ of convex domains in E_n satisfies both assumptions of Helly's Theorem 2. Indeed, if $\alpha \leqslant \xi_1 < \xi_2 < \ldots < \xi_n \leqslant \beta$, then $\Delta = D_{\xi_1} D_{\xi_2}, \ldots, D_{\xi_n}$ is evidently non-void and bounded; in fact Δ is a proper parallelepiped, except in the case $\rho_0 = 0$ when Δ reduces to a point. Let us now consider $n + 1$ distinct abscissae

$$(23) \qquad \alpha \leqslant x_0 < x_1 < \ldots < x_n \leqslant \beta.$$

If $P(x)$ is the polynomial of best approximation to the points $(x_\nu, f(x_\nu))$ ($\nu = 0, \ldots, n$), we have by (21)

$$|f(x_\nu) - a_0x_\nu^{n-1} - \ldots - a_{n-1}| = \rho(x_0,x_1, \ldots, x_n) \leqslant \rho_0 \qquad (\nu = 0, \ldots, n).$$

Geometrically this means that the $n + 1$ convex domains $D_{x_0}, D_{x_1}, \ldots, D_{x_n}$ have the common point $(a_0, \ldots, a_{n-1})$. By Helly's theorem we conclude the existence of a point $(a^*_0, \ldots, a^*_{n-1})$ which is common to *all* the domains D_x, hence there exists a polynomial $P^*(x)$ satisfying the inequality (22) for *all* x. For this polynomial $P^*(x)$ we therefore have

$$(24) \qquad \sup_{\alpha \leqq x \leqq \beta} |f(x) - P^*(x)| \leqslant \rho_0.$$

https://doi.org/10.4153/CJM-1950-022-8 Published online by Cambridge University Press

On the other hand, for an *arbitrary* polynomial $P(x)$ we have

$$\sup_{\alpha \leqq x \leqq \beta} |f(x) - P(x) \geqslant \sup_{\nu} |f(x_\nu) - P(x_\nu)| \geqslant \rho(x_0, \ldots, x_n),$$

or

$$\sup_{\alpha \leqq x \leqq \beta} |f(x) - P(x)| \geqslant \rho(x_0, \ldots, x_n).$$

Taking the supremum of the right-hand side we find

$$\sup_{\alpha \leqq x \leqq \beta} |f(x) - P(x)| \geqslant \rho_0, \text{ for every } P(x), \tag{25}$$

in particular also

$$\sup_{\alpha \leqq x \leqq \beta} |f(x) - P^*(x)| \geqslant \rho_0. \tag{26}$$

Now by (24) and (26) we find that

$$\sup_{\alpha \leqq x \leqq \beta} |f(x) - P^*(x)| = \rho_0. \tag{27}$$

From (25) and (27) we see ρ_0 is the minimum value of (18), hence $\rho = \rho_0$, which is what we wanted to prove.

REMARKS. 1. The polynomial $P^*(x)$, whose existence has just been proved, need not be unique. Thus, if $n = 2$ and

$$f(x) = [x] \qquad (0 \leqslant x \leqslant 1), \tag{28}$$

then a graph will show that every polynomial of the family

$$P(x) = a_0(x - 1) + \tfrac{1}{2} \qquad (0 \leqslant a_0 \leqslant 1),$$

minimizes the expression (18), giving it its minimal value $\rho = \frac{1}{2}$.

2. The existence part of Theorem 4 is also easily established directly by familiar continuity arguments; however, a proof of the relation (19), which seems to be new at least for a discontinuous $f(x)$, would be difficult or at least involved without the use of Helly's theorem which bridges most naturally the gap between the finite (algebraic) best approximation problem for $n + 1$ points and the analytical problem for an interval $[\alpha,\beta]$.

3. Theorem 4 immediately generalizes to the case when the interval $[\alpha,\beta]$ of definition of $f(x)$ is replaced by an arbitrary bounded point-set of the x-axis. A further possible extension of Theorem 3 is as follows: The inequality (22) means that the curve $y = P(x)$ intersects the family of vertical segments $f(x) - \rho_0 \leqslant y \leqslant f(x) + \rho_0$, which are all of the same length $2\rho_0$. As in the Dresher-Harris theorem, this length could be required to vary with x.

7. The classical case of continuous functions. We now add the important additional assumption that the function $f(x)$ is *continuous* in the range $[\alpha,\beta]$. Then it is clear that $\rho(x_0, \ldots, x_n)$, defined by (20), is a continuous function of $(x_0, \ldots, x_n)$ as long as the inequalities (23) hold. We now extend the definition of the function $\rho(x_0, \ldots, x_n)$ throughout the closed domain

$$\alpha \leqslant x_0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant \beta, \tag{29}$$

by the convention that

$$\rho(x_0, x_1, \ldots, x_n) = 0 \quad \text{if the } x_\nu \text{ are not all different.} \tag{30}$$

https://doi.org/10.4153/CJM-1950-022-8 Published online by Cambridge University Press

We claim that the extended function $\rho(x_0, \ldots, x_n)$ is continuous throughout the closed domain (29). Indeed, let $(x_0, \ldots, x_n)$ be a point of (29) for which at least two x_ν's coalesce, hence $\rho(x_0, \ldots, x_n) = 0$. Let this point $(x_0, \ldots, x_n)$ be the limit of a sequence of points $(x_0^{(k)}, \ldots, x_n^{(k)})$ with

$$\alpha \leqslant x_0^{(k)} < x_1^{(k)} < \ldots < x_n^{(k)} \leqslant \beta \qquad (k = 1,2, \ldots).$$

We have to show that

$$\lim_{k \to \infty} \rho(x_0^{(k)}, \ldots, x_n^{(k)}) = 0. \tag{31}$$

Clearly there exists a polynomial $P(x)$, of degree $n - 1$ or less, so that

$$f(x_\nu) - P(x_\nu) = 0 \qquad (\nu = 0, \ldots, n),$$

hence by the continuity of $f(x)$ and $P(x)$

$$\max_\nu |f(x_\nu^{(k)}) - P(x_\nu^{(k)})| \to 0, \text{ as } k \to \infty.$$

This, together with the inequality

$$\max_\nu |f(x_\nu^{(k)}) - P(x_\nu^{(k)})| \geqslant \rho(x_0^{(k)}, \ldots, x_n^{(k)}),$$

implies (31).

Let us now return to our Theorem 4 to note the effect of the continuity of $f(x)$. Let us assume that the best approximation ρ is positive. The continuous function $\rho(x_0, \ldots, x_n)$ *assumes* its maximum value ρ at a point $(x^*_0, \ldots, x^*_n)$ and because ρ is positive we must have

$$\alpha \leqslant x^*_0 < x^*_1 < \ldots < x^*_n \leqslant \beta.$$

We may now readily establish contact with the classical oscillation properties of the polynomial $P^*(x)$ of best approximation[12] ρ. In the first place $P^*(x)$ is now uniquely defined: Indeed, a polynomial $P(x)$ of best approximation $\rho = \rho(x^*_0, \ldots, x^*_n)$ must satisfy the inequalities

$$|f(x^*_\nu) - P(x^*_\nu)| \leqslant \rho = \rho(x^*_0, \ldots, x^*_n) \qquad (\nu = 0, \ldots, n), \tag{32}$$

while we know from the discussion of the case of $n + 1$ points that there is only one polynomial satisfying (32); since $P^*(x)$ does satisfy (32), $P^*(x)$ is uniquely defined. Secondly, we know that the sequence

$$u^*_\nu = f(x^*_\nu) - P^*(x^*_\nu) \qquad (\nu = 0, \ldots, n),$$

has all its elements of absolute value equal to ρ and that they alternate in sign. These are the classical oscillation properties referred to above. Our example (28) shows that no such properties, beyond the general relation (19) hold in the case of discontinuous functions.

Appendix (Added July 1, 1949). The authors are much indebted to the referee for the following two valuable references;

1. The theorem of Dines and McCoy of our Introduction is an immediate corollary of a theorem of E. Steinitz, *Bedingt konvergente Reihen und konvexe Systeme* II, Journal für Mathematik, vol. 144 (1941), pp. 1-40. On pp. 12-13

[12]See [9], pp. 76-78.

Steinitz defines a family of rays in E_n, with common initial point O, to be *all-sided* provided there are rays of the family on each side of every hyperplane through O. An all-sided family is *irreducible* if no proper sub-family is all-sided. Steinitz then proves the following

THEOREM. *In any all-sided family of rays, there is contained at least one irreducible sub-family; such a sub-family has at least $n + 1$ and at most $2n$ rays.*

The Dines-McCoy theorem for the space E_n, rather than E_{n+1}, follows thus: Let $\{H_\nu\}$ be the collection of half-spaces of the theorem and let R_ν denote the interior ray through O normal to the hyperplane bounding H_ν. Suppose that these half-spaces have no common ray. Then for every ray ρ through O, for some ν we must have $\angle(R_\nu, \rho) > \pi/2$; applying this remark to ρ and $-\rho$, we see that $\{R_\nu\}$ is an all-sided family of rays. By Steinitz's theorem there is an all-sided sub-collection $R_1, R_2, \ldots, R_s$, say, with $n + 1 \leqslant s \leqslant 2n$. But then the corresponding $H_1, H_2, \ldots, H_s$ have no ray in common, in contradiction to the assumption of the theorem.

2. The Dresher-Harris theorem of our Introduction was fully discussed (for $n = 1$) by L. A. Santaló, *Complemento a la nota: Un teorema sobre conjuntos de paralelepipedos de aristas paralelas*, Publicaciones del Instituto de Matematica de la Universidad Nacional del Litoral, vol. 3 (1942), pp. 203-210. Also its proof by means of Helly's theorem is found in footnote 4 on page 207 and attributed by Santaló to J. Rey Pastor.

REFERENCES

[1] P. Alexandroff and H. Hopf, *Topologie*, vol. 1, Berlin, 1935.

[2] L. M. Blumenthal, *Metric methods in linear inequalities*, Duke Math. J., vol. 15 (1948), 955-966.

[3] L. L. Dines and N. H. McCoy, *On linear inequalities*, Trans. Roy. Soc. Can., Third Series, Sec. III, vol. 27 (1933), 37-70.

[4] E. Helly, *Über Mengen konvexer Körper mit gemeinschaftlichen Punkten*, Jahresbericht der deutschen Mathematiker Vereinigung, vol. 32 (1923), 175-176.

[5] P. Kirchberger, *Über Tschebyschefsche Annäherungsmethoden*, Math. Ann., vol. 57 (1903), 509-540. The same paper appeared also in more elaborate form (96 pages) in 1902 as a doctoral dissertation written under Hilbert's guidance.

[6] D. König, *Über konvexe Körper*, Math. Zeit., vol. 14 (1922), 208-210.

[7] J. Radon, *Mengen konvexer Körper, die einen gemeinsamen Punkt enthalten*, Math. Ann., vol. 83 (1921), 113-115.

[8] A. Tarski, *A decision method for elementary algebra and geometry*, The Rand Corporation, 1948, 57 pages.

[9] Ch. J. de la Vallée Poussin, *Leçons sur l'approximation des fonctions d'une variable réelle*, Paris, 1919.

The University of Pennsylvania