Replacing points by compacta in neural network approximation

Paul C. Kainen (Washington, DC)
To Hewitt Kenyon, in memoriam *

Abstract

It is shown that cartesian product and pointwise-sum with a fixed compact set preserve various approximation-theoretic properties. Results for pointwise-sum are proved for F-spaces and so hold for any normed linear space, while the other results hold in general metric spaces. Applications are given to approximation of L_p-functions on the d-dimensional cube, $1 \leq p < \infty$, by linear combinations of half-space characteristic functions; i.e., by Heaviside perceptron networks.

1 Introduction

Dugundji remarks that in Hausdorff spaces, "the compact subsets behave as points do and have the same separation properties" ([5, p. 225]. For example, in such spaces two disjoint compact subsets have disjoint neighborhoods. That is, compacta can replace points.

Our aim in this paper is to apply this paradigm of Dugundji to some problems which arise in nonlinear approximation. The basic idea was first applied in our work with Kůrková and Sanguineti [11] under more restrictive hypotheses (normed linear spaces only). Here we have drawn further consequences for analysis from the underlying topology and geometry.

---

* *2000 Mathematics Subject Classification*: Primary 54E35; Secondary 54B10, 41A65. J. Franklin Inst., Vol. 341/4 pp 391-399, July, 2004

Key words and phrases: distance-minimizing sequence, well-posedness, approximative compactness, proximinal set, cartesian product, pointwise-sum, F-space, nonlinear approximation, Heaviside neural network, satisficing, regularization.

It is shown that in metric spaces compacta can replace points with respect to cartesian product of subsets, and for pointwise-sum when the metric space is an F-spaces. Since F-spaces include normed linear spaces, our results apply in typical function spaces used in nonlinear approximation. Instead of separation properties, we consider a hierarchy of compactness-type properties (defined precisely below) applying to subsets of a metric space. A subset is proximinal if it contains a best approximation to any point in the metric space and a subset is approximatively compact if for every point any minimizing sequence for its distance functional restricted to the subset converges subsequentially to an element of the subset.

We show that for approximative compactness and proximinality, points can be replaced by compact sets (Theorem 3.1, 4.1, respectively); also that cartesian product (Theorem 5.1), and pointwise-sum in the F-space case (Theorem 5.2) preserve the compactness hierarchy properties when we operate by cartesian product or sum with a compact subset. Further, the set of points in an approximatively compact subset which minimize the distance to a given compact subset is itself compact (Theorem 5.4). An application is given in section 6 to neural network approximation.

The paper is organized as follows. Section 2 contains preliminaries. Section 3 generalizes results in [11] and Sections 4 and 5 contain the main results, and the last section concerns applications as in [11].

2 Preliminaries

Let (X, d) be a metric space. Kuratowski's notation [14] is used in denoting the distance between two points $x, y \in X$ by $|x-y|$ rather than $d(x, y)$. When there is no linear structure, the notation is unambiguous. In case there is a linear structure as well, we add some conditions on the metric, which are weaker than assuming it comes from a norm.

An *F-space* is a real linear space endowed with a metric which satisfies (i) *invariance of the metric* $|\, x - y \,| = |\, (x - y) - 0 \,|$ (i.e., the distance from x to y equals the distance from $x - y$ to 0); (ii) the mapping $(\alpha, x) \to \alpha\, x$ from $\Re \times X$ to X is continuous in α for each x and continuous in x for each α; and (iii) the metric space is complete [6, pp. 51-52]. For an F-space X, addition is a continuous mapping of $X \times X$ to X, and multiplication by -1 is an isometry; also, any normed linear space is an F-space. See [6, pp. 52-53].

By invariance of the metric, in an F-space, the notation $|a - b - c|$ is unambiguous since (iv) $|(a - b) - c| = |(a - b) - c - 0| = |a - (b + c)|$. Moreover, we have a corresponding result for subsets which is needed in section 5. For subsets A, B of a linear space, let $A + B$ denote the set of pairwise sums $A + B = \{a + b : a \in A, b \in B\}$; $a + B$ means $\{a\} + B$; and let $-B$ denote $\{-b : b \in B\}$.

If X is a metric space, for $A, B \subset X$, let $|A - B| = \inf_{a \in A,\ b \in B} |a - b|$ and for $a \in X$, write $|a - B|$ instead of $|\{a\} - B|$. The following lemma is derived from (iv) by taking infima.

Lemma 2.1 *Let X be an F-space with nonempty subsets A and B. For any $x \in X$, $|A + B - x| = |A - (x - B)|$.*

A sequence t_n *converges to* t_0 if for every $\varepsilon > 0$, there exists N such that for all $n > N$, $|t_n - t_0| < \varepsilon$. A sequence *converges subsequentially* if it has a convergent subsequence; our notation

$$t_n \geq t_{n'} \longrightarrow t_0$$

identifies the subsequence and the point to which it converges. Recall that a subset C of a metric space is compact if every sequence in C converges subsequentially to an element of C; a subset is *conditionally compact* if it has compact closure with respect to the metric-induced topology. Also, given sequences s_n, t_n, and a subsequence $s_{n'}$ of the first sequence, the corresponding subsequence of the second is denoted $t_{n'}$. Compact sets are called *compacta*. Compactness equals countable compactness for metric spaces [9, p. 133].

A subset of a metric space is *boundedly compact* if every bounded sequence in the subset is subsequentially convergent. In our notation above, Y is boundedly compact if for any bounded sequence y_n in Y, there is a point x_0 (not necessarily in Y) for which $y_n \geq y_{n'} \longrightarrow x_0$. A well-known theorem of Riesz asserts that a normed linear space is boundedly compact if and only if it is finite-dimensional [2, p. 40].

For a metric space X and nonempty subsets M and C, we say that a sequence $m_n \in M$ *converges in distance* to C if $\lim_{n \to \infty} |m_n - C| = |M - C|$; the subset M is *approximatively compact relative to* C if every sequence $m_n \in M$ which converges in distance to C is subsequentially convergent to an element of M. This is equivalent to the generalized Tykhonov well-posedness of the problem of minimizing the distance to C from M [4, p. 24].

Call M a subset of X *approximatively compact* provided that M is approximatively compact relative to each of the singletons of X; M is *proximinal* (or an "existence set") if for every $x \in X$ some element m in M satisfies the equation $|x-m| = |x-M|$. These notions arose in nonlinear approximation theory [7],[19],[3].

Akhieser shows that a closed subspace of Hilbert space is approximatively compact [1, p.24], but it follows from the theorem of Riesz noted above that such a subspace is not boundedly compact. We give a nonlinear example of the same phenomenon in section 6.

There is a *compactness hierarchy*: compact, closed and boundedly compact, approximatively compact, proximinal, closed. Each of these properties implies the next [18, pp. 382-384].

3 Distance minimization

The theorems in this section were originally obtained for normed linear spaces in [11]. The first result says that points can be replaced by compact subsets in the definition of approximative compactness.

Theorem 3.1 *Let M and C be nonempty subsets of a metric space X. If M is approximatively compact and C is compact, then M is approximatively compact relative to C.*

Proof. Let $m_n \in M$ be any sequence converging in distance to C and let the sequence $c_n \in C$ satisfy

$$(*) \quad \lim_{n\to\infty} |m_n - c_n| = |M - C|.$$

Since C is compact, $c_n \geq c_{n'} \to c_0 \in C$. Hence, for every $\varepsilon > 0$ there exists N such that for $n' > N$, $|M - C| \leq |m_{n'} - c_0| \leq |m_{n'} - c_{n'}| + |c_{n'} - c_0| \leq |M - C| + \varepsilon$. Therefore, $m_{n'}$ converges in distance to c_0 so, since M is approximatively compact, $m_n \geq m_{n'} \to m_0 \in M$; that is, m_n converges subsequentially to an element of M. □

All we need for the theorem is that M is approximatively compact relative to all singletons in the closure of C and that C is conditionally compact. By adding conditions on M, we can reduce the requirements for C as is shown in the next two results.

Theorem 3.2 *Let M and C be nonempty subsets of a metric space X. If M is approximatively compact and bounded, and C is boundedly compact, then M is approximatively compact relative to C.*

Proof. Let $m_n \in M$ be any sequence converges in distance to C and let $c_n \in C$ satisfy $(*)$. As m_n is bounded, so is c_n. Since C is boundedly compact, $c_n \geq c_{n'} \to c_0 \in X$. Proceed as in the proof of Theorem 3.1. $\square$

Again, it is only necessary to assume that M is approximatively compact relative to singletons in the closure of C.

Theorem 3.3 *Let M and C be nonempty subsets of a metric space X. If M is closed and boundedly compact and C is bounded, then M is approximatively compact relative to C.*

Proof. Suppose m_n converges in distance to C and again choose c_n in C such that $(*)$ holds. As c_n is bounded, so is m_n; hence, $m_n \geq m_{n'} \to m_0 \in M$. $\square$

4 Distance realization

Distance between a pair of compact sets can be realized by a pair of points in the respective sets [16], [9, p. 141]. This can be generalized.

Theorem 4.1 *Let M and C be nonempty subsets of a metric space X. If M is proximinal and C is compact, then there exist points $m \in M$ and $c \in C$ with $|m - c| = |M - C|$.*

Proof. Suppose $c_n \in C$ satisfies $\lim_{n\to\infty} |M - c_n| = |M - C|$. By compactness of C, $c_n \geq c_{n'} \to c_0 \in C$ so $|M - c_0| = |M - C|$. Now choose $m_0 \in M$ such that $|m_0 - c_0| = |M - c_0|$. $\square$

Theorem 4.2 *Let M and C be nonempty subsets of a metric space X. If M is proximinal and bounded and C is closed and boundedly compact, then there exist points $m \in M$ and $c \in C$ with $|m - c| = |M - C|$.*

Proof. Suppose $c_n \in C$ satisfies $\lim_{n\to\infty} |M - c_n| = |M - C|$. Since M is bounded, c_n must also be bounded so $c_n \geq c_{n'} \to c_0 \in C$. Proceed as in the proof of Theorem 4.1. $\square$

In both of these theorems, M need only contain best approximations to elements from C. Also, distance realization is not always possible when M and C are closed and boundedly compact [11]. In the Euclidean plane, take the real line and the graph of $y = e^x$. Hence, [18, Theorem 2.3, p. 385] is incorrect.

5 Operations preserving compactness

Let X and Y be metric spaces. We give the cartesian product $X \times Y$ the metric $|(x, y) - (x', y')| = |x - x'| + |y - y'|$. The topology induced on $X \times Y$ is the product topology (e.g., [9, p. 62]).

The next result implies that cartesian product with a compact set preserves the compactness hierarchy.

Theorem 5.1 *Let S and P be nonempty subsets of metric spaces X and Y, respectively. Suppose that P is compact. If S is boundedly compact or approximatively compact, then so is $S \times P$.*

Proof. If S is boundedly compact, we show that any sequence (s_n, p_n) in $S \times P$ which is bounded has a convergent subsequence. Indeed, by definition of the product metric, s_n is bounded and since S is boundedly compact, $s_n \geq s_{n'} \to s_0 \in X$. By compactness of P $p_{n'} \geq p_{n''} \to p_0 \in P$. Hence, $(s_n, p_n) \geq (s_{n''}, p_{n''}) \to (s_0, p_0) \in X \times Y$.

If S is approximatively compact, let (x, y) be any element in $X \times Y$ and suppose that (s_n, p_n) is a sequence in $S \times P$ which converges in distance to (x, y); that is, $\lim\limits_{n\to\infty} |(s_n, p_n) - (x, y)| = |S \times P - (x, y)|$. By compactness of P, $p_n \geq p_{n'} \to p_0 \in P$. Hence, $\lim\limits_{n'\to\infty} |(s_{n'}, p_0) - (x, y)| = |S \times P - (x, y)|$ so $\lim\limits_{n'\to\infty} (|s_{n'} - x| + |p_0 - y|) = |S - x| + |P - y|$. Thus, the limit $L = \lim\limits_{n'\to\infty} |s_{n'} - x|$ exists and $L = |S - x| + |P - y| - |p_0 - y| \leq |S - x|$. But $|s - x| \geq |S - x|$ for all $s \in S$ so $L = |S - x|$. Hence, $s_{n'}$ converges in distance to x and since S is approximatively compact, $s_{n'} \geq s_{n''} \to s_0 \in S$. Therefore, $(s_n, p_n) \geq (s_{n''}, p_{n''}) \to (s_0, p_0) \in S \times P$; i.e., $S \times P$ is approximatively compact $\square$

The next theorem shows that adding a compact subset in an F-space also preserves the compactness hierarchy.

Theorem 5.2 *Let S and P be nonempty subsets of an F-space X. Suppose that P is compact. If S is boundedly compact, approximatively compact, proximinal, or closed, respectively, then so is $S + P$.*

Proof. Let S be boundedly compact and suppose $s_n + p_n$ is a bounded sequence in $S + P$. By compactness of P, p_n is bounded, so s_n is bounded and hence $s_n \geq s_{n'} \to s_0 \in X$. Again by compactness of P, any sequence in it contains a convergent subsequence; i.e., $p_{n'} \geq p_{n''} \to p_0 \in P$. Hence, $s_n + p_n \geq s_{n''} + p_{n''} \to s_0 + p_0$. Therefore, $S + P$ is boundedly compact.

For preservation of approximative compactness, we use previous results. Let S be approximatively compact and let $s_n + p_n$ be a sequence in $S + P$ converging in distance to $x \in X$; i.e., $\lim_{n\to\infty} |s_n + p_n - x| = |S + P - x|$. As P is compact, $p_n \geq p_{n'} \to p_0 \in P$. Hence, using Lemma 2.1, $\lim_{n'\to\infty} |s_{n'} - (x - p_0)| = \lim_{n'\to\infty} |s_{n'} + p_0 - x| = |S + P - x| = |S - (x - P)|$. Therefore, $s_{n'}$ converges in distance to the compact set $x - P$. Thus, by Theorem 3.1, $s_{n'} \geq s_{n''} \to s_0 \in S$; hence, $S + P$ is approximatively compact.

Let S be proximinal and let $x \in X$. Then $x - P$ is compact so by Theorem 4.1, there exist elements $s_0 \in S$ and $x - p_0 \in x - P$ for which $|s_0 - (x - p_0)| = |S - (x - P)|$. Hence, by Lemma 2.1, $|s_0 + p_0 - x| = |S + P - x|$.

For a proof that S closed implies $S + P$ closed, see Holmes [10, p. 6]. □

As with respect to distance, the attributes of compactness used in arguments can be redistributed (cf. Theorem 3.2).

Theorem 5.3 *Let S, P be nonempty subsets of an F-space X. If S is approximatively compact and bounded and P is closed and boundedly compact, then $S + P$ is approximatively compact.*

The proof is immediate using Lemma 2.1 and Theorem 3.2. Also, one may check that in Theorems 5.1 and 5.2, bounded compactness is preserved provided only that P is conditionally compact. Also, if $\mu : G \times G \to G$ is a topological group (Hausdorff) with C compact and M closed, then the set MC of products $\mu(m, c)$, $m \in M$, $c \in C$, is closed [6, p. 414].

Another operation preserving compactness is "metric projection" (see, e.g., [18]) to an approximatively compact subset.

Theorem 5.4 *Let M and C be nonempty subsets of a metric space X. If M is approximatively compact and C is compact, then $K = \{m \in M : \exists c \in C, |m - c| = |M - c|\}$ is compact.*

Proof. Let y_n be a sequence in K and for every n choose c_n in C so that y_n minimizes the distance from M to c_n. Since C is compact, $c_n \geq c_{n'} \to c_0 \in C$. Hence, for every $\varepsilon > 0$, there exists N such that for all $n' > N$, $|c_{n'} - c_0| < \varepsilon$; therefore, for all $n' > N$,

$$|M - c_0| \leq |y_{n'} - c_0| \leq |y_{n'} - c_{n'}| + |c_{n'} - c_0| = |M - c_{n'}| + |c_{n'} - c_0| < |M - c_0| + 2\varepsilon.$$

Therefore, $y_{n'}$ converges in distance to c_0, so it converges subsequentially. □

It follows that $\{m \in M : |m - C| = |M - C|\}$ is compact when C is compact and M is approximatively compact. Thus, in a metric space, the metric projection of a compact subset into an approximatively compact subset is compact.

6 Application to Heaviside neural nets

An n-fold linear combination of half-space characteristic functions corresponds exactly to a feedforward perceptron network with n hidden units, each having the Heaviside threshold function as its activation function. For $X = L_p([0,1]^d)$, d a positive integer and $1 \leq p < \infty$, we showed in [12], [13] that the nonconvex subset $M = span_n H_d$ consisting of all n-fold linear combinations of half-space characteristic functions restricted to the d-cube is approximatively compact for every positive integer n.

However, for $n \geq 2$, $span_n H_d$ cannot be boundedly compact in the metric space X since it contains a unit-norm sequence. Indeed, let κ_j be the half-space characteristic function $x_1 \leq 1/j$ restricted to the d-dimensional cube and let κ_0 be the characteristic function of the cube itself. Put $g_j = j^{1/p}(\kappa_0 - \kappa_j)$. Then the p-norm of g_j is 1 for all positive integers j but g_j cannot converge to a bounded function.

The Dugundji paradigm described above applies to nonlinear approximation by replacing a "target function" with a compact target family. Given input-output data, only known to be approximately correct, and some bound on smoothness, there will be a compact family of functions which agree with the input-output data up to a given measurement tolerance and which also possess at least the required degree of smoothness. (This is a consequence of the Sobolev embedding theorems.) For details, see [11]. Note, however, that this family could be empty if the data is not compatible with the required smoothness.

Given the data including a bound on smoothness, one seeks to find an n-fold linear combination of half-space characteristic functions with minimum distance to the compact set C consisting of all functions satisfying the smoothness and data constraints. The theorems given above and in [11] guarantee that such best approximants do exist provided that C is nonempty. Moreover, these best approximants occur as the limit of a subsequence of any sequence which minimizes the distance functional of C restricted to M.

It is the author's conjecture that approximative compactness also holds with any compact convex subset replacing the unit cube. Additional approximatively compact subsets can be constructed using the above theorems via pointwise-sum and cartesian product with compacta.

Proximinality yields the existence of a function implementable by an n-hidden-node Heaviside-perceptron neural network which achieves the minimum possible distance (in the L_p-sense) to any function agreeing with the data up to the given tolerance and meeting the smoothness requirement. This approach might be termed "satisficing" in H. Simon's sense. Moreover, approximative compactness guarantees the convergence of a greedy algorithm aimed at finding this function. Also, our method yields nonunique optimum solutions, which allows for additional control.

In contrast, Tykhonov regularization minimizes a weighted combination of fit to data and smoothness. See, e.g., [17]. It obtains a unique minimum solution which depends on the relative weight assigned to the smoothness (or other regularizing) functional. While there is a "representer theorem" providing through matrix inversion a unique solution, expressed in terms of a linear combination based on the given data and regularization parameter, the size of the data set is often too large for this to be practical.

Regularization can be combined with our approach as well [15].

In engineering applications, it would be interesting to interpret the notion of *disturbance* (see, e.g., [8]) using cartesian product or pointwise-sum with compact or boundedly compact subsets.

References

[1] N. I. Akhieser, *Theory of Approximation*, Dover, New York, 1992 (orig. 1956).

[2] M. M. Day, *Normed Linear Spaces*, Academic Press, New York, 1962.

[3] F. Deutsch, *Existence of best approximations*, J. of Approximation Theory 28 (1980), 132–154.

[4] A. L. Dontchev and T. Zolezzi, *Well-Posed Optimization Problems*, Lecture Notes in Math., vol. 1543. Springer-Verlag, Berlin Heidelberg, 1993.

[5] J. Dugundji, *Topology*, Allyn and Bacon, Boston, 1966.

[6] N. Dunford and J. T. Schwartz, *Linear Operators*, Part I, Interscience Publ., New York, 1967.

[7] N. V. Efimov and S. B. Stečkin, *Approximative compactness and Čebyshev sets*, Dokl. Akad. Nauk SSSR 140 (1961) 522-524 (Russian); English transl. in Soviet Math. Dokl. 2 (1961).

[8] R. J. P. de Figuierdo and G. Chen, *Optimal disturbance rejection for nonlinear control systems*, IEEE Trans. on Automatic Control 34 (1989), 1242–1248.

[9] S. A. Gaal, *Point Set Topology*, Academic Press, New York, 1964.

[10] R. B. Holmes, *A Course on Optimization and Best Approximation*, Lecture Notes in Mathematics 257, Springer-Verlag, Berlin, 1972.

[11] P. C. Kainen, V. Kůrková, and M. Sanguineti, *Minimization of error functionals over variable basis functions*, SIAM J. of Optimization 14 (2003), 732–742.

[12] P. C. Kainen, V. Kůrková, and A. Vogt, *Best approximation by Heaviside perceptron networks*, Neural Networks 13 (2000), 695–697.

[13] P. C. Kainen, V. Kůrková, and A. Vogt, *Best approximation by linear combinations of characteristic functions of half-spaces*, J. of Approximation Theory 122 (2003), 151–159.

[14] C. Kuratowski, *Topology*, vol. I, Academic Press, New York, 1966.

[15] V. Kůrková and M. Sanguineti, Learning with generalization capability by kernel methods of bounded complexity. Research report ICS-2003-901, submitted for publication.

[16] M. Nicolescu, *Sur la meillieure approximation d'une fonction donnée par les fonctions d'une famille donnée*, Bul. Fac. sti. Cernauti 12 (1938), 120–128.

[17] T. Poggio and S. Smale, *The mathematics of learning: dealing with data*, Notices of the American Math. Soc. 50 (2003), 536–544.

[18] I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces*, Springer-Verlag, Berlin, 1970.

[19] L. P. Vlasov, *The concept of approximative compactness and its variants*, Mat. Zametki 16 (1974), 337–348; English transl. Math Notes 16 (1974) 786–792.

Department of Mathematics
Georgetown University
Washington, DC 20057 U.S.A.
kainen@georgetown.edu