

**MATH491**

**Summer Research Project**

**2006– 2007**

Constructive Approximation Theory

**Matthew Hendtlass**

Department of Mathematics and Statistics  
University of Canterbury

# Constructive Approximation Theory

Matthew Hendtlass

February 9, 2007

## Abstract

We consider the fundamental Theorem of Approximation Theory from a constructive viewpoint, revealing that the theorem itself is fundamentally non-constructive. We subsequently present a development of a constructive alternative to the Fundamental Theorem, under the hypothesis that there is at most one best approximation in our linear space. Basic applications of this theorem are discussed and brief mention made regarding alternative additions to the hypothesis of the Fundamental Theorem.

# 1 Introduction

Bishop's constructive mathematics provides a framework in which all results can in theory be implemented on a computer. It is not the only algorithmic alternative to classical mathematics, but it has the advantage that, although the proofs may seem complicated, all proofs within that framework are classically valid.

The key way in which Bishop's constructive mathematics differs from classical mathematics is in the assertion, of the former, that an object exists if and only if it can be constructed. This means that we must reinterpret the logical connectives and quantifiers so that any proof of existence involves a construction and any decisions involved can be made by a finite intelligence (for example, computers or people). Arend Heyting provided an axiomatic formal system that captured the principles used in constructive reasoning. The underlying **BHK-interpretation** of the logical connectives and quantifiers is the following:

- $P \vee Q$ : we have either a proof of  $P$  or a proof of  $Q$ .
- $P \wedge Q$ : we have a proof of  $P$  and a proof of  $Q$ .
- $P \Rightarrow Q$ : we can convert any proof of  $P$  into a proof of  $Q$ .
- $\neg P$ : assuming  $P$ , we can derive a contradiction.
- $\exists x P(x)$ : we have an algorithm that computes a certain  $x$  and another that shows that  $P(x)$  holds.
- $\forall x \in A P(x)$ : we have an algorithm which, applied to  $x$  and a proof that  $x \in A$ , shows that  $P(x)$  holds.

This interpretation leads to certain important classical equivalences failing to hold. First, it is obvious that  $P \vee Q$  implies  $\neg(\neg P \vee \neg Q)$ . Under traditional, or "classical", logic, the converse also holds; that is, in order to prove  $P \vee Q$  it is sufficient to prove that it is impossible for both  $P$  and  $Q$  not to hold. While intuitively this seems reasonable, working constructively it is clear that  $\neg(\neg P \vee \neg Q)$  does not provide sufficient information to decide (prove) whether  $P$  holds or  $Q$  holds, and so constructively

$$\neg(\neg P \vee \neg Q) \not\Rightarrow P \vee Q.$$

An even more important consequence of our reinterpretation of the logical connectives and quantifiers results in the rejection of proof by contradiction, which relies on the classical equivalence

$$\exists P(x) \leftrightarrow \neg \forall x \neg P(x).$$

Working constructively, we see that the impossibility that  $P(x)$  fails to hold for all  $x$  in no way allows us to construct an  $x$  for which  $P(x)$  holds. The other direction of implication is, however, still valid.

These two fundamental differences between classical and constructive mathematics can be regarded as a result of the rejection of the **Law of Excluded Middle (LEM)**, which says that for all statements  $P$ ,

$$P \vee \neg P$$

holds. Classical mathematics, in fact, is equivalent to Bishop's constructive mathematics plus LEM.

Other "omniscience principles", weaker than the full LEM, which we must reject to ensure our mathematics is constructive include:

- **The Limited Principle of Omniscience (LPO):** For each binary sequence  $(a_n)_{n \geq 1}$ , either  $a_n = 0$  for all  $n$  or else there exists an  $n$  such that  $a_n = 1$ .
- **The Lesser Principle of Omniscience (LLPO):** For each binary sequence  $(a_n)_{n \geq 1}$  with at most one term equal to 1, either  $a_n = 0$  for all even  $n$  or  $a_n = 0$  for all odd  $n$ .
- **The Axiom of Choice:** Let  $C$  be a collection of inhabited<sup>1</sup> sets. Then there exists a so-called "choice function"  $f$  defined on  $C$  with the property that, for each set  $S$  in  $C$ ,  $f(C)$  is a member of  $S$ .
- **Markov's Principle:** For any binary sequence  $(a_n)_{n \geq 1}$ , if it is impossible for all terms to be equal to 0, then there exists a term that equals 1.

The Axiom of Choice may not at first seem nonconstructive, but it was shown by Diaconescu [8] in a categorical setting, and later, in the context of set theory, by Goodman-Myhill [9], to imply LEM. There are, however, choice axioms that are often (but not universally) accepted in constructive mathematics; namely the **Principle of Countable Choice**, which states that any countable collection of non-empty sets has a choice function, and the weaker **Principle of Dependent Choice**:

If  $X$  is a set,  $a \in X$ ,  $S$  is a subset of  $X \times X$ , and for each  $x \in X$  there exists  $y \in X$  such that  $(x, y) \in S$ , then there exists a sequence  $(x_n)_{n \geq 1}$  in  $X$  such that  $x_1 = a$  and  $(x_n, x_{n+1}) \in S$  for each positive integer  $n$ .

Markov's Principle likewise is perhaps not obviously nonconstructive. It is rejected because it represents an unbounded search.

We must be careful in interpreting the statement  $x \neq 0$  for real numbers  $x$ . Constructively we define

$$x \neq 0 \Leftrightarrow |x| > 0.$$

Classically this is equivalent to the standard definition:  $x \neq 0$  if and only  $\neg(x = 0)$ . But constructively, the negative notion  $\neg(x = 0)$  is computationally

---

<sup>1</sup>A set  $S$  is said to be **inhabited** if there exists an element of  $S$ ; this is a stronger property than the impossibility of  $S$  being empty. In fact, the equivalence of these two properties implies LEM.

weaker than the positive one  $|x| > 0$ . In fact, the correspondence of these two notions of  $x \neq 0$  is equivalent to Markov's Principle. Rejecting that principle as nonconstructive, we use the computationally stronger condition  $|x - y| > 0$  as our defining condition for the inequality relation  $\neq$  on the set of real numbers.

The exclusion of the omniscience principles from constructive mathematics has serious constructive consequences. For instance, though the statement

$$\forall x (x = 0 \vee x \neq 0) \tag{1}$$

might seem unproblematic, it fails to hold constructively. To show this we present a "Brouwerian Counterexample": an example showing that a classically valid proposition constructively entails a nonconstructive principle such as the omniscience principles outlined above.

Let  $(a_n)_{n \geq 1}$  be a binary sequence, and consider the binary number

$$x = \sum_{n=1}^{\infty} 2^{-n} a_n.$$

Assuming that the statement (1) holds under the constructive interpretation, we can decide either that  $x = 0$  or that  $|x| > 0$ . In the first case we must have  $a_n = 0$  for all  $n$ . In the second we can compute a positive integer  $N$  such that  $|x| > 2^{-N}$ . If  $a_n = 0$  for all  $n \leq N$ , then

$$x = \sum_{n=N+1}^{\infty} 2^{-n} a_n \leq \sum_{n=N+1}^{\infty} 2^{-n} = 2^{-N},$$

a contradiction. Since we can decide, for each  $n$ , whether  $a_n = 0$  or  $a_n = 1$ , by testing  $a_1, \dots, a_N$  in turn, we are guaranteed to find  $n \leq N$  such that  $a_n = 1$ . Hence for each binary sequence  $(a_n)_{n \geq 1}$  either  $a_n = 0$  for all  $n$  or else there exists an  $n$  such that  $a_n = 1$ . In other words, (1) implies LPO and is therefore nonconstructive.

Another example of a classically "obvious" statement that fails to hold constructively is

$$\forall x (x \geq 0 \vee x \leq 0). \tag{2}$$

This can be shown to imply LLPO by letting  $(a_n)_{n \geq 1}$  be a binary sequence with at most one term equal to 1, and considering the alternating binary expansion

$$x = \sum_{n=1}^{\infty} (-1)^n 2^{-n} a_n.$$

Fortunately there are useful constructive alternatives to such essentially omniscient statements as (1) and (2). Examples are the **cotransitivity law**,

$$\forall x \in \mathbb{R} \forall y \in \mathbb{R} \forall z \in \mathbb{R} (x < y \Rightarrow x < z \vee z < y),$$

and the initially surprising statement

$$\forall x \in \mathbb{R} (\neg(x > 0) \Rightarrow x \leq 0).$$

The surprise is elicited by the fact that the statement

$$\forall x \in \mathbb{R} (\neg(x \geq 0) \Rightarrow x < 0)$$

implies LPO.

For further information on constructive mathematics see [6] or [1], to which the reader is directed for justification of statements that appear below with neither proof nor associated reference.

## 2 Existence of Best Approximations

Our main aim in this report is to explore constructively the foundations of approximation theory.

Let  $X$  be a metric space,  $V$  a subspace of  $X$ , and  $x \in X$ . A **best approximation** to  $x$  in  $V$  is an element  $v_0 \in V$  such that  $\rho(x, v_0) \leq \rho(x, v)$  for each  $v \in V$ . We say that the subspace  $V$  is **proximal** (in  $X$ ) if each  $x \in X$  has a best approximation to  $x$  in  $V$ ; in that case,  $V$  is **located**, in the sense that for each  $x \in X$  the distance

$$\rho(x, V) = \inf \{ \rho(x, y) : y \in V \}$$

exists. (Note that the statement “Every inhabited subset of  $\mathbb{R}$  is located” is equivalent to LPO.)

**Theorem 1** *All finite dimensional subspaces of a real normed space are proximal.*

Here is a classical proof. Fix  $x$  in  $X$ , and consider the closed ball  $B = \bar{B}(0, 2\|x\|)$  in  $V$ . Since  $V$  is finite dimensional, this ball is compact. But the distance function on  $X$  is (uniformly) continuous, so

$$m = \inf \{ \rho(x, s) : s \in B \} = \rho(x, B)$$

exists. For each  $y$  in  $V \setminus B$  we have  $\|y\| > 2\|x\|$  and therefore

$$\|y - x\| \geq \|y\| - \|x\| > \|x\| = \|x - 0\| \geq \rho(x, B),$$

since  $0 \in B$ . Hence  $\rho(x, V)$  exists and equals  $\rho(x, B)$ . Since  $B$  is compact and the distance from  $B$  is a continuous function,  $\rho(x, V) = \rho(x, B)$  is attained; that is, there exists  $v \in B \subset V$  such that  $\rho(x, V) = \rho(x, B)$ .

How does this proof fare constructively? The first problem we encounter is that of deciding whether  $s \in B$  or  $s \in V \setminus B$ , which clearly entails determining that either  $s \in B$  or  $s \notin B$ . But for real numbers, the statement

$$\forall s (s \in B \vee s \notin B)$$

implies LEM. However, this difficulty is easily overcome as follows.

Take

$$B = \bar{B}(0, 2\|x\| + 1)$$

For each  $y$  in  $V$  we have either  $\|y\| > 2\|x\|$  or else  $\|y\| < 2\|x\| + 1$ . In the first case,  $\|x - y\| > \|x\| \geq \rho(x, B)$  as in the foregoing classical proof. Hence  $\rho(x, V)$  exists and equals  $\rho(x, B)$ .

The real problem from a constructive standpoint is the idea that since  $B$  is compact,  $\rho(x, B)$  is attained. For there is no algorithm that, applied to any uniformly continuous function on a compact interval of the line, will enable us to compute a point where that function attains its infimum; indeed, the existence of such an algorithm is equivalent to LLPO (8.3.2 of [10]). So we see that our classical proof of the Fundamental Theorem, with the minor adjustment given in the second to last paragraph, provides only a constructive proof that  $\rho(x, B)$  exists. In other words, we have a constructive proof that all finite-dimensional subspaces of a normed space are located.

So our classical proof of the Fundamental Theorem is essentially nonconstructive. But is this just a failure of this particular proof or is it indicative of a more fundamental failure? This question is answered by the following Brouwerian example (due to F. Richman).

**Proposition 2** *The Fundamental Theorem of Approximation theory implies LLPO.*

**Proof.** Let  $X = \mathbb{R}^2$  with the maximum norm

$$\|(x, y)\| = \max\{|x|, |y|\},$$

and let

$$V = \mathbb{R}(\cos \theta, \sin \theta),$$

where  $\theta \in \mathbb{R}$ . Given  $|\theta| \ll \pi/2$ , suppose that we can find a best approximation, say  $(x, y)$ , to the point  $(0, 1)$  in  $V$ . We may assume that  $|\theta|$  is so small that  $|x| > |y|$ . Either  $x < 1/2$  or  $x > -1/2$ . In the first case, if  $\theta > 0$  then  $x > 0$  and

$$\begin{aligned} \rho((0, 1), (x, y)) &= \max\{|x|, |1 - y|\} \\ &= 1 - y \\ &> 1 - y' \\ &= \rho((1, 0), (1/2, y')) \end{aligned}$$

where  $(1/2, y') \in \mathbb{R}(\cos \theta, \sin \theta)$ . This contradicts that  $(x, y)$  is a best approximation to  $(0, 1)$ . Hence  $\neg(\theta > 0)$  and so (by Lemma 2.1.4. of [6])  $\theta \leq 0$ . Similarly if  $x > -1/2$  then  $\theta \geq 0$  and so the Fundamental Theorem implies that for all real  $\theta$ , either  $\theta \geq 0$  or  $\theta \leq 0$ . This, in turn, implies LLPO.

■

**Corollary 3** *The Fundamental Theorem of Approximation theory is constructively equivalent to LLPO.*

**Proof.** Since the statement

*Every uniformly continuous mapping of a compact set into the real line attains its infimum*

is equivalent to LLPO, the desired conclusion follows directly from the preceding proposition. ■

Thus we are forced to conclude that a new "fundamental theorem" is required, rather than a new proof.

Examining Richman's Brouwerian counterexample, we see that the choices for  $X$  and  $V$  are essentially the prototypical examples of a vector space and a finite-dimensional linear space and subsequently the most fundamental choices of  $X$  and  $V$ . So if we hope to come up with a reasonable constructive alternative to the Fundamental Theorem, we must assume that the problem lies elsewhere.

The use of the maximum norm in Richman's Brouwerian example is essential: if the same example is considered with the Euclidean norm instead, then the best approximation to  $(0, 1)$  exists. Furthermore, the obvious place where problems arise is when  $\theta$  is close to 0; in fact, taking  $\theta = 0$  and using the maximum norm, we see that there are infinitely many best approximations to  $(0, 1)$ —namely, the elements of the set  $\{(x, 0) : x \in [0, 1]\}$ .

So we have to add something to the hypotheses of the Fundamental Theorem in order to avoid this particular counterexample. The "something" introduced in [2] is given in the next definition.

**Definition.** Let  $X$  be a normed space,  $V$  an  $n$ -dimensional subspace of  $X$ , and  $a$  an element of  $X$ . We say that  $a$  has at most one best approximation in  $V$  if

$$\max\{\|a - v\|, \|a - v'\|\} > \rho(a, V)$$

whenever  $v, v'$  are distinct elements of  $V$ .

We now present, without proof, a few results (found in [6] as 2.2.14, 4.2.2, and 4.2.3 respectively) that we shall call upon later.

**Proposition 4** *Let  $f$  be a uniformly continuous mapping of a totally bounded metric space  $X$  into  $\mathbb{R}$ . Then the set*

$$X(f, r) = \{x \in X : f(x) \leq r\}$$

*is either compact or empty for all but countably many  $r \in \mathbb{R}$ .*

**Lemma 5** *Let  $\{e_1, \dots, e_n\}$  be a basis of an  $n$ -dimensional subspace  $V$  of a normed space  $X$ , let  $1 \leq m < n$ , and let  $W$  be the subspace of  $X$  with basis  $\{e_1, \dots, e_m\}$ . Then the span of  $\{e_{m+1}, \dots, e_n\}$  is an  $(n - m)$ -dimensional subspace of the quotient space  $X/W$ .*

**Lemma 6** *Let  $x, e$  be elements of a real normed space  $X$  with  $e \neq 0$ , and for each  $\delta > \rho(x, \mathbb{R}e)$  write*

$$S_\delta = \{t \in \mathbb{R} : \|x - te\| < \delta\}.$$

*If  $S_\delta$  is compact, then it is a proper compact interval  $[m, M]$  in  $\mathbb{R}$ . Moreover,*

$$\|x - me\| = \delta = \|x - Me\|.$$

Without being able to make the omniscient choices embodied by principles such as LLPO it is not surprising that, in proving a constructive version of the Fundamental Theorem, we (must?) resort to using induction on the dimension of our finite dimensional subspace. To this end we present the following lemma, which provides the basis for this induction.

**Lemma 7** *Let  $X$  be a real normed space, let  $x, e$  be points of  $X$  with  $e \neq 0$ , and let  $d \geq 0$ . If for all distinct  $t, t' \in \mathbb{V}$  we have  $\max\{\|x - te\|, \|x - t'e\|\} > d$ , then there exists  $\tau \in \mathbb{R}$  such that*

$$\|x - \tau e\| > d \Rightarrow \rho(x, \mathbb{R}e) > d.$$

In order to prove this lemma, we use a very powerful constructive technique, the “ $\lambda$ -technique”, which, typically under the assumption that one of the spaces involved is complete, allows us to avoid otherwise necessary appeals to omniscience principles. Here we mark the attainment of one of two alternatives, dependent on  $n$ , by the values of a binary sequence  $(\lambda_n)_{n \geq 1}$ . In most cases, if  $\lambda_n = 1$  the desired property is attained. We then construct an appropriate Cauchy sequence in our complete space, and using its limit, find  $N$  such that  $\neg(\lambda_n = 0)$ ; in that case we must have  $\lambda_n = 1$ , so the desired property holds. For more on this, see Chapter 3 of [6].

This technique often leads to proofs of statements that at first seem nonconstructive, such as the preceding lemma, which we now prove.

**Proof.** The basic idea is to create a sequence  $(a_n)_{n \geq 1}$  such that the terms  $\rho(x, a_n)$  are successively closer to  $\rho(x, \mathbb{R}e)$ . For each  $n$  we make the decision that either  $\rho(x, \mathbb{R}e) < d + 1/n$  or  $\rho(x, \mathbb{R}e) > d$ , letting  $\tau = a_N$  if  $N$  is the smallest integer for which we have decided that  $\rho(x, \mathbb{R}e) > d$ . In that case,

$$\neg(\rho(x, \mathbb{R}e) < d + 1/N),$$

so  $\rho(x, \mathbb{R}e) > d$ . We then proceed to show that such an  $N$  exists.

To formalise this we fix  $e \in \mathbb{R} \setminus \{0\}$  and make use of the closed convex sets

$$S_{\delta_n} = \{t \in \mathbb{R} : \|x - te\| \leq \delta_n\}$$

introduced in Lemma 6. We construct an increasing binary sequence  $(\lambda_n)_{n \geq 1}$ , real sequences  $(a_n)_{n \geq 1}$  and  $(b_n)_{n \geq 1}$ , and a strictly decreasing sequence  $(\delta_n)_{n \geq 1}$  of positive numbers, such that:

- $S_{\delta_n}$  is the proper compact interval  $[a_n, b_n]$ ;
- if  $\lambda_n = 0$ , then  $\rho(x, \mathbb{R}e) < \delta_n < 1 + 1/n$  and, if  $n \geq 2$ ,

$$0 < b_n - a_n \leq \frac{2}{3} (b_{n-1} - a_{n-1}); \quad (3)$$

- if  $\lambda_n = 1$ , then  $\rho(x, \mathbb{R}e) > d$ .

To start the construction, we note that either  $\rho(x, \mathbb{R}) > d$  or  $\rho(x, \mathbb{R}) < d + 1$ . In the former case there is nothing to prove, so without loss of generality we can assume  $\rho(x, \mathbb{R}) < d + 1$  and set  $\lambda_n = 0$ . Using Proposition 4 and Lemma 6, we then compute  $\delta_1 > 0$  such that  $S_{\delta_1}$  is a proper compact interval  $[a_1, b_1]$ . Next, suppose that we have constructed  $\lambda_{n-1}$ ,  $a_{n-1}$ ,  $b_{n-1}$ , and  $\delta_{n-1}$  with the desired property. If  $\lambda_{n-1} = 1$ , set

$$\lambda_n = 1, \quad a_n = a_{n-1}, \quad b_n = b_{n-1}, \quad \delta_n = \delta_{n-1}. \quad (4)$$

If  $\lambda_{n-1} = 0$ , let

$$\begin{aligned} \varepsilon &= \frac{1}{6} (b_{n-1} - a_{n-1}), \\ c &= \frac{1}{2} (a_{n-1} + b_{n-1}). \end{aligned}$$

By our hypothesis,

$$r \equiv \max\{\|a - (c - \varepsilon)e\|, \|a - (c + \varepsilon)e\|\} > d,$$

so either  $\rho(x, \mathbb{R}e) > d$  or else  $\rho(x, \mathbb{R}e) < \min\{r, d + \frac{1}{n}\}$ . In the first case we define our numbers as at 4. In the second we set  $\lambda_n = 0$  and, using Proposition 4, choose  $\delta_n$  such that

$$\rho(x, \mathbb{R}e) < \delta_n < \min\left\{\delta_{n-1}, r, d + \frac{1}{n}\right\}$$

and  $S_{\delta_n}$  is compact. By Lemma 6,  $S_{\delta_n}$  is a proper compact interval  $[a_n, b_n]$  contained in  $S_{\delta_{n-1}}$ .

Since  $\max\{\|a - (c - \varepsilon)e\|, \|a - (c + \varepsilon)e\|\} > d$  either  $\|a - (c - \varepsilon)e\| > d$  or  $\|a - (c + \varepsilon)e\| > d$ . In the first case  $a - (c - \varepsilon)e \notin S_{\delta_n}$  and so, since  $S_{\delta_n}$  is convex,  $S_{\delta_n} \subset [a_{n-1}, c - \varepsilon e]$  or  $S_{\delta_n} \subset (c - \varepsilon e, b_{n-1}]$ . In the second  $S_{\delta_n} \subset [a_{n-1}, c + \varepsilon e]$  or  $S_{\delta_n} \subset (c + \varepsilon e, b_{n-1}]$ . In all cases,

$$b_n - a_n = |S_{\delta_n}| \leq \frac{2}{3} (b_{n-1} - a_{n-1}),$$

our inductive construction is complete.

From 10, for  $\lambda_n = 0$  and  $n \geq 2$  we now have

$$0 < b_n - a_n \leq \left(\frac{2}{3}\right)^{n-1} (b_1 - a_1)$$

and thus, in light of 4, for all  $n$ ,

$$0 \leq a_{n-1} - a_n \leq \left(\frac{2}{3}\right)^{n-1} (b_1 - a_1)$$

since, by virtue of the construction process,  $a_{n+1} \geq a_n$  and  $b_n \geq a_{n+1}$ . Hence  $(a_n)_{n \geq 1}$  is a Cauchy sequence and therefore converges to a limit  $t \in \mathbb{R}$ .

Finally, suppose that  $\|x - te\| > d$ , and compute  $N$  such that

$$\|x - a_N e\| > d + \frac{1}{N}.$$

If  $\lambda_N = 0$ , then

$$\|x - a_N e\| = \delta_N > d + \frac{1}{N},$$

a contradiction. Hence  $\lambda_N = 1$ , and therefore  $\rho(x, \mathbb{R}e) > d$ . This completes the proof of Lemma 7. ■

**Definition.** A subspace  $V$  of a metric space  $X$  is said to be **quasiproximinal** (in  $X$ ) if for each  $x \in X$  that has at most one best approximation in  $V$ , there exists a best approximation to  $x$  in  $V$ .

We now present a constructive version of the Fundamental Theorem of Approximation Theory.

**Theorem 8** *Every finite-dimensional subspace of a real normed space is quasiproximinal.*

**Proof.** Let  $V$  be a finite-dimensional subspace of a real normed space  $X$ , and let  $a \in X$  have at most one best approximation in  $V$ . If  $\dim(V) = 0$ , then the result is trivial. If  $\dim(V) = 1$ , then we pick  $e \neq 0$  such that  $V = \mathbb{R}e$ , and apply Lemma 7 with  $x = a$  and  $d = \rho(a, V)$  to construct  $t \in \mathbb{R}$  such that  $\|a - te\| = \rho(a, V)$ . Then  $te$  is a best approximation to  $x$  in  $V$ . We complete the proof by induction on  $\dim(V)$ . Accordingly, suppose we have proved that all  $n$ -dimensional subspaces of real normed spaces are quasiproximinal, and consider the case where  $V$  is  $n$ -dimensional, with basis  $\{e_1, \dots, e_{n+1}\}$ . Letting  $Y = \mathbb{R}e_{n+1}$ , we see from Lemma 5 that  $V/Y$  is an  $(n-1)$ -dimensional subspace of  $X/Y$  with basis  $\{e_1, \dots, e_n\}$ . Note that for each  $x$  in  $X$  we have

$$\rho(x, V) = \inf \left\{ \|x - v\|_{X/Y} : v \in V \right\} = \rho_{X/Y}(x, V),$$

where  $\rho_{X/Y}$  denotes the metric associated with the quotient norm on  $X/Y$ . Given  $v \in V$ , we compute  $\alpha \in \mathbb{R}$  such that if  $\|a - v - \alpha e_{n+1}\| > \rho(a, V)$ , then

$$\|a - v\|_{X/Y} = \rho(a - v, \mathbb{R}e_{n+1}) > \rho(a, V).$$

To do so, note that for distinct real numbers  $t, t'$  we have

$$\|(v + te_{n+1}) - (v + t'e_{n+1})\| = |t - t'| \|e_{n+1}\| \neq 0,$$

so, by our hypothesis that  $\mathbf{a}$  has at most one best approximation in  $V$ ,

$$\max\{\|\mathbf{a} - v - te_{n+1}\|, \|\mathbf{a} - v - t'e_{n+1}\|\} > \rho(\mathbf{a}, V).$$

We can now apply Lemma 7 with  $x = \mathbf{a} - v$ ,  $e = e_n$ , and  $d = \rho(\mathbf{a}, V)$  to compute the  $\alpha$  we wanted. Having done so, let  $v' \in V$  be distinct from  $v$ , and compute, in the same way,  $\alpha' \in \mathbb{R}$  such that if  $\|\mathbf{a} - v' - \alpha'e_{n+1}\| > \rho(\mathbf{a}, V)$ , then  $\|\mathbf{a} - v'\|_{X/Y} > \rho(\mathbf{a}, V)$ . Since

$$\begin{aligned} \|(v + \alpha e_{n+1}) - (v' + \alpha' e_{n+1})\| &= \|(v - v') + (\alpha - \alpha') e_{n+1}\| \\ &\geq \|v - v'\|_{X/Y} > 0, \end{aligned}$$

our hypotheses ensure that

$$\max\{\|\mathbf{a} - v - \alpha e_{n+1}\|, \|\mathbf{a} - v' - \alpha' e_{n+1}\|\} > \rho(\mathbf{a}, V).$$

Hence, by our choice of  $\alpha$  and  $\alpha'$ ,

$$\max\{\|\mathbf{a} - v\|_{X/Y}, \|\mathbf{a} - v'\|_{X/Y}\} > \rho(\mathbf{a}, V) = \rho_{X/Y}(\mathbf{a}, V).$$

Thus  $\mathbf{a}$  has at most one best approximation in the  $n$ -dimensional subspace  $V/Y$  of the quotient space  $X/Y$ . Applying the induction hypothesis, we obtain a best approximation  $v$  to  $\mathbf{a}$  in  $V/Y$ . Then

$$\rho(\mathbf{a} - v, \mathbb{R}e_{n+1}) = \|\mathbf{a} - v\|_{X/Y} = \rho_{X/Y}(\mathbf{a}, V) = \rho(\mathbf{a}, V).$$

For this choice of  $v$ , construct  $\alpha$  (as earlier) such that if  $\|\mathbf{a} - v - \alpha e_{n+1}\| > \rho(\mathbf{a}, V)$ , then  $\|\mathbf{a} - v\|_{X/Y} > \rho(\mathbf{a}, V)$ . We clearly have

$$\neg(\|\mathbf{a} - v - \alpha e_{n+1}\| > \rho(\mathbf{a}, V))$$

and therefore

$$\|\mathbf{a} - v - \alpha e_{n+1}\| = \rho(\mathbf{a}, V).$$

This completes the inductive proof of our theorem. ■

Theorem 8 is classically equivalent to the classical Fundamental Theorem. For if no best approximations to  $x$  in  $V$  exist, then clearly  $x$  has at most one best approximation in  $V$ , and therefore, by the theorem, it has a best approximation in  $V$ —a contradiction.

For an application of Theorem 8 we introduce a property that applies to many important types of normed space (such as the  $L_p$ -spaces—see Chapter 6 of [1]).

**Definition.** A normed space  $X$  is said to be **uniformly convex** if for each  $\varepsilon > 0$  there exists  $\delta$  with  $0 < \delta < 1$  such that if  $x, y$  are unit vectors in  $X$  and  $\|\frac{1}{2}(x + y)\| > 1 - \delta$ , then  $\|x - y\| < \varepsilon$ .

Thus  $X$  is uniformly convex if, when the average of two unit vectors is nearly a unit vector, the two unit vectors are close. Uniform convexity rules out corners in balls of the space. (It is easily seen that  $\mathbb{R}^2$ , taken with the maximum norm, is not uniformly convex, precisely because its unit ball has sharp corners.)

**Proposition 9** *Every finite-dimensional subspace of a real uniformly convex normed space is proximal.*

**Proof.** Let  $X$  be a real uniformly convex normed space,  $V$  a finite-dimensional subspace of  $X$ , and  $a$  a point of  $X$ . In view of Theorem 8, it is enough to prove that  $a$  has at most one best approximation in  $V$ . To this end let  $v, v' \in V$  be distinct and let  $\alpha = \|v - v'\| > 0$ . Set

$$\varepsilon = \frac{\alpha}{3(1 + \rho(a, V))}$$

and choose  $\delta \in (0, 1)$  such that

$$\rho(a, V) + \frac{\alpha}{6} < \frac{\alpha}{6(1 - \delta)}$$

and such that if  $x, y \in V$  are unit vectors with  $\|\frac{1}{2}(x + y)\| > 1 - \delta$ , then  $\|x - y\| < \varepsilon$ . For convenience, let

$$m = \max\{\|a - v\|, \|a - v'\|\}.$$

Either  $\rho(a, V) < m$  or  $m < \rho(a, V) + \alpha/6$ . In the latter case, if  $\rho(a, V) < \alpha/6$ , then

$$\begin{aligned} \alpha = \|v - v'\| &\leq \|a - v\| + \|a - v'\| \\ &< 2\left(\rho(a, V) + \frac{\alpha}{6}\right) < \alpha \end{aligned}$$

a contradiction; so  $\rho(a, V) \geq \alpha/6 > 0$ , and therefore both  $\|a - v\| \geq \alpha/6$  and  $\|a - v'\| \geq \alpha/6$ . Defining

$$\begin{aligned} w &= \frac{1}{\|a - v\|}(a - v), \\ w' &= \frac{1}{\|a - v'\|}(a - v'), \end{aligned}$$

we have

$$\begin{aligned} \|w + w'\| &= \left\| \frac{1}{\|a - v\|}(a - v) + \frac{1}{\|a - v'\|}(a - v') \right\| \\ &= \left\| \left( \frac{1}{\|a - v\|} + \frac{1}{\|a - v'\|} \right) a - \left( \frac{1}{\|a - v\|}v + \frac{1}{\|a - v'\|}v' \right) \right\| \\ &= \left( \frac{1}{\|a - v\|} + \frac{1}{\|a - v'\|} \right) \|a - v''\|, \end{aligned}$$

where

$$v'' = \frac{\|a - v\| \|a - v'\|}{\|a - v\| + \|a - v'\|} \left( \frac{1}{\|a - v\|} v + \frac{1}{\|a - v'\|} v' \right) \in V.$$

Hence

$$\begin{aligned} \|w + w'\| &\geq \left( \frac{1}{\|a - v\|} + \frac{1}{\|a - v'\|} \right) \rho(a, V) \\ &\geq \frac{2}{\rho(a, V) + \alpha/6} \times \frac{\alpha}{6} \\ &> \frac{6(1 - \delta)}{\alpha} \times \frac{\alpha}{3} \\ &= 2(1 - \delta) \end{aligned}$$

and therefore  $\|\frac{1}{2}(w + w')\| > 1 - \delta$ . Since  $w, w'$  are unit vectors, it follows that  $\|w - w'\| < \varepsilon$ . Thus

$$\begin{aligned} \|v - v'\| &= \left\| \left( 1 - \frac{\rho(a, V)}{\|a - v\|} \right) (v - a) - \rho(a, V)(w - w') \right. \\ &\quad \left. + \left( 1 - \frac{\rho(a, V)}{\|a - v'\|} \right) (a - v') \right\| \\ &\leq \|a - v\| - \rho(a, V) \|w\| + \rho(a, V) \|w - w'\| \\ &\quad + \|a - v'\| - \rho(a, V) \|w'\| \\ &= \|a - v\| - \rho(a, V) + \rho(a, V) \|w - w'\| + \|a - v'\| - \rho(a, V) \\ &< \frac{\alpha}{6} + \frac{\alpha \rho(a, V)}{3(1 + \rho(a, V))} + \frac{\alpha}{6} \\ &< \alpha. \end{aligned}$$

Once again we have arrived at a contradiction to the definition of  $\alpha$ . This rules out the case  $m < \rho(a, V) + \alpha/6$ . Hence  $m > \rho(a, V)$ , so  $a$  has at most one best approximation in  $V$ . ■

Theorem 8 says that any finite-dimensional subspace of a real normed space  $X$  with at most one best approximation to  $x \in X$  contains one; that is we can construct a best approximation to  $x$  in  $V$ . The hypothesis then ensures that this best approximation is unique. In light of this, it seems natural that one should attempt to apply Theorem 8 to situations where best approximations are classically unique, perhaps by converting a classical proof of the uniqueness of a best approximation into a constructive proof that there is at most one best approximation. This is precisely what we have done in Proposition 9.

### 3 Chebyshev Approximation

An important case where best approximations are classically unique is Chebyshev approximation over  $[0, 1]$ . In this case, given elements  $\phi_1, \phi_2, \dots, \phi_n$  of the

space  $C[0, 1]$  of continuous functions on  $[0, 1]$  taken with the sup norm

$$\|\phi\| = \sup \{\phi(x) : x \in [0, 1]\},$$

we define

$$\Phi(x) = (\phi_1(x), \dots, \phi_n(x)) \quad (x \in [0, 1])$$

and

$$\|\Phi\| = \sup \{\|\Phi(x)\|_2 : x \in [0, 1]\},$$

where  $\|\cdot\|_2$  is the usual Euclidean norm on  $\mathbb{R}^n$ . Now define mappings  $\beta, \gamma : (0, 1/n] \rightarrow \mathbb{R}^+$  as follows. For each  $\alpha \in (0, 1/n]$ ,

- if  $n = 1$ , then

$$\beta(\alpha) = \inf \{|\phi_1| : x \in [0, 1]\};$$

- if  $n \geq 2$ , then

$$\beta(\alpha) = \inf \left\{ |\det[\phi_j(x_i)]| : 0 \leq x_1, \dots, x_n \leq 1, \min_{1 \leq i < j \leq n} |x_i - x_j| \geq \alpha \right\}$$

and

$$\gamma(\alpha) = \min \left\{ \|\Phi\|, \frac{\beta(\alpha)}{n^{1/2}(n-1)! \prod_{i=1}^n (1 + \|\Phi\|)} \right\}.$$

Note that in the case  $n \geq 2$ ,  $\beta(\alpha)$  is well defined since the mapping

$$(x_1, \dots, x_n) \rightsquigarrow |\det[\phi_j(x_i)]|$$

is uniformly continuous on the compact set

$$\left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : 0 \leq x_1, \dots, x_n \leq 1, \min_{1 \leq i < j \leq n} |x_i - x_j| \geq \alpha \right\}.$$

**Definition.** We call the set  $\{\phi_1, \dots, \phi_n\}$  of elements of  $C[0, 1]$  a **Chebyshev system** if

$$\beta(\alpha) > 0 \quad \text{for each } \alpha \in (0, 1/n].$$

This condition is classically equivalent to the **Haar condition**: every finite linear combination of  $\phi_1, \dots, \phi_n$  has at most  $n - 1$  zeroes in  $[0, 1]$ .

The best known, and perhaps most important, example of a Chebyshev system is the set of polynomial functions  $\{1, x, x^2, \dots, x^n\}$  over  $[0, 1]$ .

If  $\{\phi_1, \dots, \phi_n\}$  is a Chebyshev system, then the functions  $\phi_1, \dots, \phi_n$  are linearly independent and span an  $n$ -dimensional subspace  $H$  of  $C[0, 1]$ . Given  $a \in C[0, 1]$  and  $p \in H$ , we say that an ordered pair  $(j, (x_1, \dots, x_n))$ , with  $j \in \{0, 1\}$  and  $0 \leq x_1 < x_2 < \dots < x_{n+1} \leq 1$ , is an **alternant** of  $a$  and  $p$  if

$$(-1)^{k-j} (a - p)(x_k) = \|a - p\| \quad (1 \leq k \leq n + 1).$$

Classically,  $p$  is a best approximation to  $a$  in  $H$  if and only if there exists an alternant of  $a$  and  $p$ . Since this characterisation of “best Chebyshev approximations” fails constructively (see [3]), in the constructive development of the theory we replace the notion of alternant by the following weaker one.

For  $\varepsilon > 0$ , and  $\varepsilon$ -alternant of  $a \in C[0, 1]$  and  $p \in H$  is an ordered pair  $(j, (x_1, \dots, x_n))$  such that  $j \in \{0, 1\}$ ,  $0 \leq x_1 < x_2 < \dots < x_{n+1} \leq 1$ , and

$$(-1)^{k-j} (a - p)(x_k) > \|a - p\| - \varepsilon \quad (1 \leq k \leq n + 1).$$

We then have this characterisation of best Chebyshev approximations.

**Theorem 10** *Let  $H$  be the linear span of a Chebyshev system in  $C[0, 1]$ , and let  $a \in C[0, 1]$ . A necessary and sufficient condition for  $b \in H$  to be a best Chebyshev approximation to  $a$  in  $H$  is that, for each  $\varepsilon > 0$ , there exists an  $\varepsilon$ -alternant of  $a$  and  $b$  ([5], Theorem 2).*

With the help of Theorem 10, the constructive existence of best Chebyshev approximations can be established as a special case of Theorem 8; see [3]. However, the constructive attainment of a best Chebyshev approximation can be established more directly, using Theorem 10 but without appeal to Theorem 8. This direct approach actually provides more numerical information than the former one, which is hardly surprising since Theorem 8 is a very general one.

There is a classical algorithm for finding best Chebyshev approximations: the Remes algorithm. It is interesting to note though, that the classical proof of the convergence of the Remes algorithm uses a contradiction argument, and consequently provides no information on its rate of convergence. For numerical analysts this is hardly an ideal situation, and highlights the value of developing approximation theory constructively. In the case of a slightly modified version of the Remes algorithm, the convergence was proved (constructively), and hence rates of convergence given explicitly, by Bridges in [4].

## 4 Concluding Remarks

To end the report, we make some remarks about possible alternatives to the “at most one best approximation” condition that we used in Theorem 8. One alternative is the condition that  $a \in X$  has **isolated best approximations** in  $V$ : this means that for each  $v \in V$ , each  $e \neq 0$  in  $V$ , and each  $\varepsilon > 0$ , there exists  $\lambda \in \mathbb{R}$  such that  $0 < |\lambda| < \varepsilon$  and  $\|a - v - \lambda e\| > \rho(e, V)$ . However, this condition is actually equivalent to  $a$  having at most one best approximation in  $V$ ! To see this, first observe that if  $a$  has at most one best approximation, then, almost trivially, it has isolated best approximations. Conversely, if it has isolated best approximations, consider distinct points  $v, v'$  of  $V$ . Set  $e = v' - v$  and pick a scalar  $\lambda$  such that  $0 < |\lambda| < 1/2$  and

$$\left\| a - \left( v + \frac{1}{2}e \right) - \lambda e \right\| > \rho(a, V).$$

Then, with  $0 < \alpha = 1/2 + \lambda < 1$  we have

$$\begin{aligned}
(1 - \alpha) \rho(\mathbf{a}, V) + \alpha \rho(\mathbf{a}, V) &= \rho(\mathbf{a}, V) \\
&< \left\| \mathbf{a} - \mathbf{v} - \left( \frac{1}{2} + \lambda \right) (\mathbf{v}' - \mathbf{v}) \right\| \\
&= \left\| \mathbf{a} - \mathbf{v} - \alpha (\mathbf{v}' - \mathbf{v}) \right\| \\
&= \left\| (1 - \alpha) (\mathbf{a} - \mathbf{v}) + \alpha (\mathbf{a} - \mathbf{v}') \right\| \\
&\leq (1 - \alpha) \|\mathbf{a} - \mathbf{v}\| + \alpha \|\mathbf{a} - \mathbf{v}'\|,
\end{aligned}$$

so either  $\|\mathbf{a} - \mathbf{v}\| > \rho(\mathbf{a}, V)$  or  $\|\mathbf{a} - \mathbf{v}'\| > \rho(\mathbf{a}, V)$ .

Another possible condition is that best approximations to  $\mathbf{a}$  in  $V$  are isolated relative to a given basis  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  of  $V$ , in the sense that for each  $\mathbf{v} \in V$ , each  $i$  ( $1 \leq i \leq n$ ), and each  $\varepsilon > 0$ , there exists  $\lambda \in \mathbb{R}$  such that  $|\lambda| < \varepsilon$  and  $\|\mathbf{a} - \mathbf{v} - \lambda \mathbf{e}_i\| > \rho(\mathbf{a}, V)$ . However, the adaptation of the proof of Theorem 8 that one requires in order to produce a corresponding result with this new condition breaks down at the induction step: since the quotient norm is at most the original norm on  $X$ , there seems no hope of proving that best approximations, relative to the quotient norm, in the subspace of  $X/Y$  with basis  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  are isolated relative to that basis. In fact, the sought-for result does not hold, as the following Brouwerian example shows:

Let  $X = \mathbb{R}^3$ , with the norm

$$\|(x, y, z)\| = \max\{|x|, |y|\} + |z|,$$

and let

$$V = \text{span}\{(\cos \theta, \sin \theta, 1), (0, 0, 1)\}$$

for some fixed  $\theta \in \mathbb{R}$ . With  $\mathbf{a} = (0, 1, 0)$ , it is easy to show that the preceding paragraph holds and that the best approximation to  $\mathbf{a}$  is in the subspace  $\mathbb{R}(\cos \theta, \sin \theta, 0)$  of  $V$  (that is, when  $z = 0$ ). This gives us essentially the same counterexample we had for the classical Fundamental Theorem.

**Acknowledgements.** Thanks must go to the University of Canterbury for their financial generosity and to Prof. Douglas Bridges for his generosity of time and expertise.

## References

- [1] E.A. Bishop and D.S. Bridges, *Constructive Analysis*, Grundlehren der math. Wissenschaften **279**, Springer-Verlag, Heidelberg-Berlin-New York, 1985.
- [2] D.S. Bridges, 'On the Foundations of Best Approximation Theory', *J. Approx. Theory* **28**(3), 273-292, 1980.

- [3] D.S Bridges, 'A Constructive Development of Chebyshev Approximation Theory', *J. Approx. Theory* **30**(2), 99-120, 1981.
- [4] D.S Bridges, 'A Constructive Analysis of the Remes Algorithm', *J. Approx. Theory* **32**(4), 257-270, 1981.
- [5] D.S Bridges, 'Recent Progress in Constructive Approximation Theory', in: *The L.E.J Brouwer Centenary Symposium*, North-Holland, Amsterdam, 41-50, 1982.
- [6] D.S. Bridges and L.S. Viță, *Techniques of Constructive Mathematics*, Universitext, Springer New York, 2006.
- [7] E.W. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, New York, 1966.
- [8] R. Diaconescu, 'Axiom of Choice and Completion', *Proc. Amer. Math. Soc.* **51**, 176-178, 1975.
- [9] N.D. Goodman and J. Myhill, 'Choice Implies Excluded Middle', *Zeit. Math. Logik and Grundlagen der Math.* **24**, 461, 1975.
- [10] A.S. Troelstra, *Principles of Intuitionism*, Lecture Notes in Mathematics No. 95, Springer-Verlag, Berlin, 1969.