Pacific Journal of Mathematics

THE FUNDAMENTAL THEOREM OF ALGEBRA: A CONSTRUCTIVE DEVELOPMENT WITHOUT CHOICE

Fred Richman

Volume 196 No. 1 November 2000

PACIFIC JOURNAL OF MATHEMATICS
Vol. 196, No. 1, 2000

THE FUNDAMENTAL THEOREM OF ALGEBRA: A CONSTRUCTIVE DEVELOPMENT WITHOUT CHOICE

Fred Richman

Is it reasonable to do constructive mathematics without the axiom of countable choice? Serious schools of constructive mathematics all assume it one way or another, but the arguments for it are not compelling. The fundamental theorem of algebra will serve as an example of where countable choice comes into play and how to proceed in its absence. Along the way, a notion of a complete metric space, suitable for a choiceless environment, is developed.

By *constructive mathematics* I mean, essentially, mathematics that is developed along the lines proposed by Errett Bishop [1]. More precisely, I mean mathematics that is done in the context of intuitionistic logic — without the law of excluded middle. My reasons for identifying these notions are discussed in [9] and [10], the basic contention being that constructive mathematics has the same subject matter as classical mathematics.

Ruitenburg [11] treated the fundamental theorem of algebra in a choiceless environment. He proved the theorem for real numbers that are defined by Cauchy sequences of rational numbers. An awkward feature of restricting to this class of real numbers is that, absent countable choice, Cauchy sequences of real numbers need not converge — the space is not sequentially complete. This seems a little bizarre: We introduce the real numbers because the rational numbers are not complete, but then neither are the real numbers! I would suggest that a proper treatment of the real numbers should not be based on Cauchy sequences — that the essence of a real number is that it can be approximated by rational numbers, not that it is a limit of a *sequence* of rational numbers.

In this paper I develop a general theory of completions of metric spaces that is suited to operating without countable choice. I also take a different view of what it means to construct the spectrum of a polynomial — its set of roots. Such a shift is essential because we can't even solve quadratics, in the customary sense, without countable choice. The idea is that the spectrum of a polynomial of degree n is an element of the completion of the set of n-multisets of complex numbers in an appropriate metric. Such an element manifests itself by producing, for each $\varepsilon > 0$, an n-multiset of complex numbers $\{r_1, \ldots, r_n\}$ so that $\prod_{i=1}^{n}(X - r_i)$ is within ε of the given

monic polynomial $p(X)$. I believe that this notion captures the real meaning of being able to calculate the roots of $p(X)$.

It's a little trickier to find the roots of an arbitrary nonconstant polynomial of formal degree n. The set of roots of such a polynomial naturally lie on the Riemann sphere, the complex projective line. That is, we want to construct the spectrum of the corresponding homogeneous polynomial $Y^n p(X/Y)$. In the last section I indicate how to do that.

1. Countable choice.

Let $\mathbf{N}$ denote the natural numbers. For X a set and S a subset of $X \times \mathbf{N}$, consider the following two statements.

(1) For all $n \in \mathbf{N}$ there exists $x \in X$ such that $(x, n) \in S$.
(2) There exists a sequence of elements $x_n \in X$ such that $(x_n, n) \in S$ for all $n \in \mathbf{N}$.

Clearly (2) implies (1). The axiom of *countable choice* says that (1) implies (2). Thus the axiom of countable choice asserts the existence of certain sequences in X.

This axiom is accepted as valid by all serious schools of constructive mathematics. Basically the justification comes down to analyzing the meaning of (1). Here is an argument by intuitionists [**14**, p. 189] adapted to our notation: "A proof of (1) should provide us with a method to find, for each $n \in \mathbf{N}$, an $x \in X$ such that $(x, n) \in S$; such a method is nothing else but the description of a function assigning the required x to n."

Why doesn't this argument justify the full axiom of choice in which we replace $\mathbf{N}$ by an arbitrary set Y? The difference is that the result of the procedure may depend on the way the element y in Y is presented. For example, for each real number y there is an integer m such that $y < m$, but the procedure for passing from y to m is not a function on the real numbers—indeed this would violate Brouwer's principle that every function on the real numbers is continuous, and would entail a countable form of the law of excluded middle. We can get a function from regular Cauchy sequences of rational numbers, but this function may assign different integers to sequences that represent the same real number. Presumably that problem does not arise with natural numbers because they have canonical presentations. But Lebesgue said, in a letter to Borel [**8**], "I agree completely with Hadamard when he states that to speak of an infinity of choices without giving a rule presents a difficulty that is just as great whether or not the infinity is denumerable." I am inclined to go along with Hadamard and Lebesgue—the arguments against choice do not seem to disappear when one restricts to the natural numbers.

In Russian constructive mathematics, everything is a natural number. A function from a subset S of $\mathbf{N}$ to $\mathbf{N}$ is given by a partial recursive function

that converges on S, and possibly elsewhere. Partial recursive functions can be coded as natural numbers — the Russians actually use Markov algorithms rather than partial recursive functions. The expression "for all $x \in X$ there is $y \in Y$ such that $P(x, y)$" is taken to be synonymous with "there is a partial recursive function that converges on the elements of X, takes X into Y, and $P(x, f(x))$ holds for all $x \in X$." See [**6**, p. 19]. In this setting, if X is the set of all natural numbers with the usual equality, then the function f is total and provides the choice function.

Bishop also seems to be thinking of a more structured model of this kind. The role of the partial recursive functions in Russian constructive mathematics is played by Bishop's notion of an *operation*, which can be thought of as a function from presentations to presentations. In [**1**] he says, "choice is implied by the very meaning of existence." I take this to be referring to the presentation level — Bishop was very aware of the distinction between a function that is merely onto, and one that admits a cross section (a choice function), calling the former a *surjection* and only the latter *onto*.

I want to argue here for dropping appeals to countable choice in constructive mathematics. Among "appeals to countable choice" I include any immediate passage from an instance of (1) to the corresponding instance of (2). On this view, one may appeal to countable choice without considering oneself to be doing that. When Bishop [**1**, p. 26] defines what it means for a sequence $\{x_n\}$ of real numbers to converge to a real number x_0, he says "if for each k in $\mathbf{Z}^+$ there exists N_k in $\mathbf{Z}^+$ with $|x_n - x_0| \leq 1/k$ for all $n \geq N_k$." The wording indicates that this is meant to be an instance of (1), but the notation implies that we may think of the N_k as forming a sequence that satisfies (2). Indeed, Bishop wants us to think that way.

I prefer a black-box metaphor for (1). I don't think it natural to consider how we might come to know (1), or how (1) might have been proved. Rather I consider (1) to mean simply that we can demand an x such that $(x, n) \in S$ for any $n \in \mathbf{N}$ that we choose. There is no guarantee that the x we get will be determined by n — it may depend on some other quantity that we are not even aware of, and the next time we ask we may get a different x for the same n. This is the minimalist interpretation of (1) — assuming more is contrary to how I actually think of (1) in practice.

In a completely computational model we can think of a black box as a library routine that is only guaranteed to return a suitable x. It's none of our business how the routine is implemented. If the code looks at an internal clock to decide what to do, that's fine. Or it can look at the status of some other memory location. All we can count on is that it will return a suitable x. In the computational model everything can be thought of as being a natural number—real numbers correspond to Gödel numbers of programs, and so on. Thus all function are partially defined functions from $\mathbf{N}$ to $\mathbf{N}$. This still allows *libraries* to contain nondeterministic or noncomputable functions —

oracles — which are the models for our black boxes. Indeed, if we do not allow black boxes, it is difficult to justify rejecting Church's thesis, which is classically false in its constructive formulation: *Every function from* $\mathbf{N}$ *to* $\mathbf{N}$ *is recursive*. The adoption of Church's thesis in constructive mathematics results in constructive proofs of theorems that are classically refutable—an undesirable state of affairs to my mind because I think of constructive mathematics as a generalization of classical mathematics.

I have long felt uncomfortable with arguments that use countable choice. Such arguments are employed in proving the Baire category theorem, the existence of a point in a compact set C whose distance to a given point p is positive only if the distance from p to C is positive, and the intermediate value theorem for real polynomials. In each case you are supposed to be constructing a definite sequence, but it looks like you are just verifying the possibility of continuing at each step. Initially I thought that my antipathy to these arguments stemmed from the fact that they rely on completeness, but now I suspect that it has to do with choice.

2. What's the problem with the FTA?

Countable choice is required just to construct a root of the polynomial X^2-a for arbitrary complex a. Consider the following very weak countable-choice principle.

> *Let A_n be a sequence of sets, each of which is either $\{0\}$ or consists of a pair of antipodal points on the unit circle. Suppose that if $A_n \neq \{0\}$, then $A_n = A_{n+1}$. Then there exists a sequence a_n such that $a_n \in A_n$ for all n.*

Do we really need countable choice to do that? Well, it's well known that we can't construct a function that chooses one point out of each pair of antipodal points on the circle. If we could, then by Brouwer's principle of continuity, the function would give a uniformly continuous function from the circle to itself that takes each pair of antipodes to one of them—but that's clearly impossible. Of course a discontinuous choice function can be constructed, using the law of excluded middle, by choosing the antipode whose angle is in $[0,\pi)$. But given that we can't make all those choices simultaneously, it seems pretty unlikely that we can make a generic one at some unknown point in a sequence, as in the above principle, without something like countable choice.

So we may tentatively assume that this principle cannot be derived without either choice or excluded middle. But the existence of a complex square root of every complex number implies this principle, as we can easily see. Let c_n be the square of any element in A_n (the squares of antipodal points are equal). Let $b_n = 0$ unless $c_m \neq 0$ for some $m \leq n$, in which case $b_n = c_m/m^2$. Then the sequence b_n converges to some complex number b.

Let $r^2 = b$. Then we can let $a_n = nr$ if $c_n \neq 0$, and 0 otherwise. Intuitively, the problem is that if we have a complex number b that is very near zero, we have no mechanism for choosing, in advance, between its two square roots should it turn out that $b \neq 0$.

What about finding complex roots of real cubic polynomials? Suppose we have a cubic polynomial with real coefficients, and we get the following roots to within one-half: 3, 1, and 1. Then we calculate some better approximations, and get, to within one-eighth: $1\frac{1}{4}$, $3\frac{1}{4}$, and $\frac{3}{4}$. Now clearly the $3\frac{1}{4}$ is the same root as the 3. But which 1 does the $1\frac{1}{4}$ go with? One is tempted to answer this with a question, or rather two questions: "Who knows?" and "Who cares?" However, this is exactly the kind of question we *must* answer if we are to calculate a specific root of the polynomial in the neighborhood of 1.

In this particular case we could resolve the ambiguity by computing the (necessarily real) root near 3 and reducing the polynomial to a real quadratic. For a real quadratic $X^2 + 2bX + c$ we can pick out a complex root r continuously by setting

$$r = \begin{cases} -b + i\sqrt{c - b^2} & \text{if } b^2 \leq c \\ \\ -b + \sqrt{b^2 - c} & \text{if } b^2 \geq c. \end{cases}$$

Although r is not defined for all pairs of real numbers (b, c), it is defined on a dense set, and is uniformly continuous on bounded sets, so it extends to all real numbers. The difference here is that the coefficients are *real*. You can construct a square root of an arbitrary real number, but not of an arbitrary complex number.

Presumably you can't do this with a general real cubic when you have three roots near zero. You might try to pick out the unique real root, or the infimum of three real roots. But it looks like you can't. If $\Re(a) < \Re(b), \Re(c)$, then a is a real root, but b and c may or may not be complex conjugates. If $\Im(a) < \Im(b)$, say $\Im(a) < 0$, then a and $\bar{a}$ are complex roots and the other is real, but you don't know where it sits.

What seems to be going on is that countable choice is the source of discontinuities — that without it you always have continuous local inverses. This is somewhat confirmed by looking at a sheaf model of the reals — modeling real numbers as continuous functions on a topological space — where countable choice fails. We can think of the coefficient a in the polynomial $X^2 - a$ as a continuous function (the identity) in a neighborhood of zero of the complex plane. Then a root of $X^2 - a$ would constitute a continuous inverse of the function z^2 in a neighborhood of zero.

The same phenomenon occurs with respect to the intermediate value theorem for real polynomials, which implies that any cubic polynomial has a

real root (see [5]). Fix a small positive number a and consider the polynomial $p_b(x) = 2x^3 + 3ax^2 + b$ which has a local maximum at $-a$ and a local minimum at 0. Try to construct a continuous function r on $(-a, a)$ so that $r(b)$ is always a root of $p_b(x)$. For $b > 0$ there is exactly one root of $p_b(x)$, and it is less than $-3a/2$. For $b = 0$ there are two roots, $-3a/2$ and 0. For $-a^2(3 - 2a) < b < 0$ there are three roots, one less than $-a$, one in $(-a, 0)$, and one greater than 0. Moreover, the one in $(-a, 0)$ is strictly increasing in b. So $r(b) < -a$ for $-a^2(3 - 2a) < b < a$. But if $b < -a^2(3 - 2a)$ there is only one root, and it is positive, so no such continuous function r exists.

An example where you do have continuous local cross sections is the theorem that for each real number x there is a natural number n with $x < n$. Clearly, for x_0 real, there is a neighborhood U of x_0 and a natural number n so that $x < n$ for all x in U.

In any event, it is not my purpose to demonstrate conclusively that countable choice is required for the fundamental theorem of algebra as normally stated. I want to see what we can prove without countable choice. In order to proceed, we have to indicate how to deal with real numbers in a choiceless environment.

3. Real, complex, and algebraic numbers.

In a choice-free development, we don't want to define real numbers to be Cauchy sequences (*Cauchy reals*). If we did that, we would need choice to prove that a Cauchy sequence of real numbers converges. The minimum we want of a real number is to be able to compute ε-approximations for each ε. This amounts to specifying located Dedekind cuts, as in [1, Chap. 2 Exer. 6] or [13, 2.2] (*Dedekind reals*).

So define a *real number* to be a nonempty subset S of the rational numbers satisfying

(1) If $x < y \in S$, then $x \in S$ (lower).
(2) There is an $x \notin S$ (bounded).
(3) If $x < y$ are rational numbers, then either $x \in S$ or $y \notin S$ (located).
(4) If $x \in S$, then there is $y > x$ also in S (open).

For example, if a_n is a Cauchy sequence of rational numbers, define S by $x \in S$ if there exists N, and rational $y > x$, such that $y < a_n$ for all $n \geq N$ (that is, x is eventually bounded below a_n).

These Dedekind cuts give an internal characterization of the sets $\{x \in \mathbf{Q} : x < r\}$, for r a real number. Alternatively we could use the sets $S_n = \{x \in \mathbf{Q} : |x - r| \leq 1/n\}$, for each positive integer n, as a model. They are nonempty, and have the property that if $x \in S_i$ and $y \in S_j$, then $|x - y| \leq 1/i + 1/j$. This approach has the virtue that it applies unchanged to any metric space, so we can complete an arbitrary metric space in the same way we pass from $\mathbf{Q}$ to $\mathbf{R}$.

Having constructed the real numbers, we construct the complex numbers in the usual way by considering $\mathbf{R}^2$. Neither of these fields is *discrete*, in the sense that given x and y, either $x = y$ or $x \neq y$. The algebraic numbers, on the other hand, form a discrete field.

An *algebraic number* is a complex number that satisfies a polynomial with rational coefficients. If E is any ring with no nilpotent elements other than 0, and no idempotents other than 0 or 1, and k is a discrete subfield of E, then the elements in E that are algebraic over k form a discrete subfield of E [7, Theorem VI.1.9]. In particular, taking k to be the rational numbers and E to be the complex numbers, the algebraic numbers form a discrete field—the extra information carried by the polynomial allows you to distinguish different elements, even though the polynomial is not involved in the definition of equality. The simplest dense, discrete subfield of the complex numbers is the field of Gaussian numbers, $\mathbf{Q}(i)$.

When we construct a complex root of a polynomial with coefficients in the algebraic numbers, we want to know that it is an algebraic number. This follows from very general algebraic considerations: *If $R \subset E \subset F$ are commutative rings, and E is integral over R (every element of E satisfies a monic polynomial with coefficients in R), then every element of F that is integral over E is also integral over R* [7, Corollary VI.1.5].

4. The FTA for algebraic numbers.

Bishop called a set $\{x_1, x_2, \ldots, x_n\}$ a *subfinite* set. Such a set differs from a finite set in that it need not be discrete — it is characterized as being an image of a finite set. Ray Mines suggested the more descriptive phrase *finitely enumerable set*. We will change the point of view a little and consider $\{x_1, x_2, \ldots, x_n\}$ to be a *finite multiset*, or an *n-multiset*. The difference is that two finite multisets $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ are *equal* if $m = n$ and there exists a permutation σ of $\{1, 2, \ldots, n\}$ such that $x_i = y_{\sigma(i)}$ for each i. The finitely enumerable sets $\{1, 1, 2\}$ and $\{1, 2, 2\}$ are equal; the finite multisets $\{1, 1, 2\}$ and $\{1, 2, 2\}$ are not.

We can define the distance between two n-multisets $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ in a metric space to be $\inf_\sigma \sup_i d(x_i, y_{\sigma(i)})$, where σ ranges over the permutations of $\{1, 2, \ldots, n\}$. The set of n-multisets in a metric space, equipped with this distance, forms a metric space. There is a problem here with equality because the distance between two multisets could be zero without their being equal. The classic example is $\{a, b\}$ and $\{a \vee b, a \wedge b\}$ in the real numbers, the problem being that proving the equality of those two sets requires the law of excluded middle. That's annoying. However if we restrict ourselves to *discrete* metric spaces—like the set of algebraic numbers—then this problem does not arise. For the general case, we must be

content that the set of n-multisets in a metric space forms a pseudo-metric space.

Let $M_n(\mathbf{A})$ be the set of n-multisets of algebraic numbers $\mathbf{A}$, and $\pi_n(\mathbf{A})$ the set of monic polynomials of degree n over the algebraic numbers. Put any of the equivalent standard metrics on $\pi_n(\mathbf{A})$, which is simply $\mathbf{A}^n$. Given an n-multiset $\{r_1, \dots, r_n\}$ in $M_n(\mathbf{A})$, we can form the polynomial $\prod_{i=1}^n (X - r_i)$ in $\pi_n(\mathbf{A})$. We want to show that this map gives a one-to-one correspondence between $M_n(\mathbf{A})$ and $\pi_n(\mathbf{A})$ that is uniformly continuous in both directions on bounded subsets.

That the map is one-to-one follows from an easy inductive argument using the fact that the algebraic numbers form a discrete field. That the map is onto is the fundamental theorem of algebra for polynomials over the algebraic numbers. Choiceless proofs of this theorem are available in the literature [7, Corollary XII.3.3, p. 296], although one must go back and verify that choice was not involved — countable choice is a blind spot for constructive mathematicians in much the same way as excluded middle is for classical mathematicians. The uniform continuity is necessary in order to give a choiceless version of the fundamental theorem of algebra for monic polynomials over the complex numbers.

How do you show that any nonconstant polynomial over a discrete subfield k of the complex numbers has a complex root? First reduce to the case where the polynomial is *separable* — relatively prime to its formal derivative. Because k is discrete, we can repeatedly use the Euclidean algorithm to write the polynomial as a product of separable polynomials, hence the polynomial itself may be assumed separable. This is exactly where things go wrong if k is not discrete: It's the possibility of indistinguishable roots that causes problems. If a separable polynomial is small at some point, then its derivative is bounded away from zero there. Thus if you can find a place where it is small enough, you can construct a root of it in the completion of k by Newton's method [7, Theorem XII.3.1, p. 295]. By considering winding numbers, you can show that it must get arbitrarily small, as Brouwer and de Loor did in [3].

Clearly the map from $M_n(\mathbf{A})$ to $\pi_n(\mathbf{A})$ is uniformly continuous on bounded subsets. To prove the uniform continuity in the other direction, we will use a simple geometric lemma, which we give in a very crude form for ease of proof.

Lemma 1. *Let P be a set of n algebraic numbers in the complex plane. For each positive rational number ρ, we can cover P with a finite number of disjoint polygons, each of diameter at most $72n\rho$, such that each point in P is at least ρ away from the boundaries of the polygons.*

Proof. Tessellate a sufficiently large portion of the complex plane with squares having Gaussian numbers as vertices and sides of length ρ. Color

red each square that contains a point of P (we can figure this out because the vertices of the squares, and the points in P, are algebraic numbers). So at most $4n$ squares are red. Now color red each square that touches a red square. Now at most $36n$ squares are red. The red squares break up into connected pieces, each of diameter at most $72n\rho$. Because we colored those touching squares red, each point in P is at least ρ away from the boundaries of the connected pieces. We can fill in any holes in the connected pieces without changing their diameters, if we want the polygons to have connected boundaries. □

Theorem 2. *The map from the set $M_n(\mathbf{A})$ of n-multisets of algebraic numbers to the set $\pi_n(\mathbf{A})$ of monic polynomials of degree n given by taking $\{r_1, r_2, \ldots, r_n\}$ to $\prod_{i=1}^{n}(X - r_i)$ is a one-to-one correspondence that is uniformly bicontinuous on bounded subsets.*

Proof. The only thing left to prove is the uniform continuity on bounded subsets for the map from $\prod_{i=1}^{n}(X - r_i)$ to $\{r_1, r_2, \ldots, r_n\}$. So suppose we are given B and $\varepsilon > 0$. We must show that there exists $\delta > 0$ such that if

$$f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 = \prod_{i=1}^{n}(X - r_i)$$

$$g(X) = X^n + b_{n-1}X^{n-1} + \cdots + b_1X + b_0 = \prod_{i=1}^{n}(X - s_i)$$

are such that $|a_i|, |b_i| < B$ and $|a_i - b_i| < \delta$, then there exists a permutation σ of $\{1, 2, \ldots, n\}$ such that $|r_i - s_{\sigma i}| < \varepsilon$ for all i.

Let

$$h_t(X) = (1 - t)f(X) + tg(X) = X^n + c_{n-1}X^{n-1} + \cdots + c_1X + c_0$$

for $0 \le t \le 1$, so $h_0 = f$ and $h_1 = g$. Also $|c_i| \le \max(|a_i|, |b_i|)$ and $|a_i - c_i| \le |a_i - b_i|$.

From the lemma we can cover $\{r_1, r_2, \ldots, r_n\}$ with a finite number of disjoint polygons, each of diameter at most ε, so that r_i is at least $\varepsilon/72n$ away from Δ, the boundaries of the polygons. Thus $|f(x)| \ge (\varepsilon/72n)^n$ for x on Δ. Choose $C \ge 1$, depending only on n, ε and B, such that $|x| \le C$ if x is on Δ. Then

$$|f(x) - h_t(x)| \le (\sup |a_i - b_i|)nC^{n-1}$$

if x is on Δ, so if we take

$$\delta = \frac{(\varepsilon/72n)^n}{2nC^{n-1}}$$

then if $\sup |a_i - b_i| < \delta$, we have $|f(x) - h_t(x)| \le (\varepsilon/72n)^n/2$ for x on Δ, so $|h_t(x)| \ge (\varepsilon/72n)^n/2$. As h provides a homotopy between f and g, which

is bounded away from zero on Δ, the winding numbers $\frac{1}{2\pi i}\oint \frac{f'(z)}{f(z)}dz$ and $\frac{1}{2\pi i}\oint \frac{g'(z)}{g(z)}dz$ are equal, where the integral is taken counterclockwise around the boundary of any of the polygons. So if a polygon contains exactly m roots of f, it must also contain exactly m roots of g. Hence we can construct a permutation σ of $\{1, 2, \dots, n\}$ such that r_i and $s_{\sigma i}$ lie in the same polygon for each i, and this σ does the trick. □

5. Completions of metric spaces.

We want to proceed from a multiset of roots of a polynomial with coefficients in the algebraic numbers, to a multiset of roots of a polynomial with coefficients in the complex numbers. Of course we know that we can't quite do that because we need not be able to construct roots of such polynomials. But while we might not be able to approximate a single root, we can approximate the entire multiset of roots — if we modify our notion of what that multiset is. The idea is to think of it as an element of the completion of the space of multisets of elements of a space X, rather than a multiset in the completion of X. We need to develop a choiceless theory of completions because the usual theory uses Cauchy sequences and doesn't even work for completing the rational numbers.

Let S be a metric space. By a *location* in S we mean a real valued function f on S with the properties

(1) $f(x) \geq |f(y) - d(x, y)|$ for all $x, y \in S$,
(2) $\inf_{x\in S} f(x) = 0$.

Note that (1) is equivalent to $d(x, y) \leq f(x)+f(y)$ and $f(y) \leq f(x)+d(x, y)$, and that it implies f is nonnegative and uniformly continuous. Moreover if $x \neq y$, that is, if $d(x, y) > 0$, then either $f(x) > 0$ or $f(y) > 0$, so f vanishes on at most one point. Every point z in S gives rise to the location f defined by $f(x) = d(x, z)$.

Note that $f(y) = \lim_{f(x)\to 0} d(x, y)$. This is immediate from the two properties of a location.

We can define a natural metric on the set $\widehat{S}$ of locations in S.

Theorem 3. *If f and g are locations on a metric space S, then*

$$d(f, g) = \sup_{y\in S} |f(y) - g(y)| = \inf_{x\in S} (f(x) + g(x))$$

exists and defines a metric on $\widehat{S}$. The natural map of S into $\widehat{S}$ is an isometry onto a dense subset of $\widehat{S}$, and every location on $\widehat{S}$ is given by a point of $\widehat{S}$.

Proof. For any x and y we have $|f(y) - d(x, y)| \leq f(x)$ and $|g(y) - d(x, y)| \leq g(x)$, so

$$|f(y) - g(y)| \leq f(x) + g(x).$$

Moreover, by choosing $x = y$ such that $g(x)$ is small, we can make the left hand side as close to the right as we please. Hence the supremum of the left hand side, and the infimum of the right hand side, exist and are equal. Thus $d(f, g)$ exists and is the metric derived of the supremum norm.

As $d(y, z) = \inf_{x \in S}(d(y, x)+d(z, x))$, the map from S into $\hat{S}$ is an isometry. If $f \in \widehat{S}$, then $|f(y) - d(x, y)| \leq f(x)$ so the distance from f to the image of x is exactly $f(x)$. As the latter can be made arbitrarily small, the image of S is dense in $\widehat{S}$. Finally suppose φ is a location on $\widehat{S}$. Then φ restricted to S is a location f on S. As φ agrees with the image of f on a dense subset of $\widehat{S}$, it must equal that image. □

We say that $\widehat{S}$ is the *completion* of S. A metric space S is *complete* if the natural map from S to $\widehat{S}$ is onto, that is, if every location on S is given by a point of S. The real numbers $\mathbf{R}$, defined as located Dedekind cuts, are complete in this sense: If f is a location on $\mathbf{R}$, then

$$r = \{q \in \mathbf{Q} : \text{there is } n \text{ such that } q < x \text{ whenever } f(x) < 1/n\}$$

is a located Dedekind cut in the rational numbers such that $f(x) = |x - r|$. (See [14, pp. 788-789] for a sheaf model where the Cauchy reals are not complete, and one where countable choice fails yet the Cauchy reals and the Dedekind reals coincide.)

As an illustration of this theory, an integrable function f on $[0, 1]$ is identified with the function $\int_0^1 |f(x) - g(x)|\, dx$ where g ranges over the uniformly continuous functions on $[0, 1]$. So the characteristic function of $[0, 1/2]$ is given by $\int_0^{1/2} |1 - g(x)|\, dx + \int_{1/2}^1 |g(x)|\, dx$, and (finite) step functions are defined similarly. What's an integrable set in general? A location in the closure of the zero-one valued step functions.

Theorem 4. *If $\varphi : A \to B$ is uniformly continuous on bounded subsets, then φ extends uniquely to a map from $\widehat{A}$ to $\widehat{B}$ that is uniformly continuous on bounded subsets. If A is a closed subset of B, and B is complete, then A is complete.*

Proof. Let f be a location on A. Define a location g on B by $g(b) = \lim_{f(a) \to 0} d(\varphi a, b)$. To see that this limit exists, note that $S = \{a : f(a) < 1\}$ is bounded, so for $a, a' \in S$, if $d(a, a')$ is small, then $d(\varphi a, \varphi a')$ is small. The result then follows from

$$\big|d(\varphi a, b) - d(\varphi a', b)\big| \leq d(\varphi a, \varphi a') \text{ and } d(a, a') \leq f(a) + f(a').$$

The inequality

$$g(b) \geq \big|g(b') - d(b, b')\big|$$

follows from the inequality $d(\varphi a, b) \geq |d(\varphi a, b') - d(b, b')|$. The map φ is unique because A is dense in $\widehat{A}$.

Now suppose that A is a subset of the complete space B, and f is a location on A. Let g be the image of f under the inclusion map of A into B. Then there exists b_0 in B such that $g(b) = d(b, b_0)$. Hence $f(a) = d(a, b_0)$. But $f(a)$ can be made arbitrarily small. So b_0 is in the closure of A, hence in A if A is closed. □

Instead of using locations to construct the completion of a metric space X, we could use *regular sequences of subsets*. A sequence of nonempty subsets S_n of X is *regular* if $d(x, y) \le 1/m + 1/n$ for each $x \in S_m$ and $y \in S_n$. Two regular sequences S and T are *equivalent* if $d(x, y) \le 2/n$ for each $x \in S_n$ and $y \in T_n$. Given a location f on X we get a (canonical) regular sequence $S_n = \{x : f(x) \le 1/n\}$. Conversely, given a regular sequence S_n, and an element x in X, then the set of $q \in \mathbf{Q}$ such that $q < d(x, y)$ for all $y \in S_m$ for sufficiently large m, defines a real number $f(x)$ as Dedekind cut, and the function f is a location.

Stolzenberg [12] suggested a similar way to specify an element ξ of the completion of a metric space X. Construct a set Σ consisting of pairs (x, c) where x is in X and c is a nonnegative real number — think of x as approximating ξ to within c. We require that there be elements (x, c) in Σ with c arbitrarily small (but not necessarily 0), and that Σ satisfy the Cauchy condition that if (x, c) and (x', c') are in Σ, then $d(x, x') \le c + c'$.

6. The FTA for monic polynomials.

We can now put the pieces together. As before, let $M_n(\mathbf{A})$ denote the set of n-multisets of algebraic numbers and $\pi_n(\mathbf{A})$ the set of monic polynomials of degree n over the algebraic numbers. The natural map from $M_n(\mathbf{A})$ to $\pi_n(\mathbf{A})$ is a one-to-one correspondence that is bicontinuous on bounded subsets. So this map extends uniquely to a one-to-one correspondence between the completions $\widehat{M}_n(\mathbf{A}) = \widehat{M}_n(\mathbf{C})$ and $\widehat{\pi}_n(\mathbf{A}) = \pi_n(\mathbf{C})$. This is the fundamental theorem of algebra for monic polynomials over the complex numbers. The space $\widehat{\pi}_n(\mathbf{A}) = \pi_n(\mathbf{C})$ is simply the set of monic polynomials of degree n with complex coefficients. What exactly is the space $\widehat{M}_n(\mathbf{A})$?

These are the *limit n-multisets* of elements of $\mathbf{A}$ (or of $\mathbf{C}$). By their approximations you shall know them. Given a monic polynomial $p(X)$ with complex coefficients — that is, an element of $\pi_n(\mathbf{C})$ — what is the corresponding element μ of $\widehat{M}_n(\mathbf{A})$? The approximations to μ are multisets $\{r_1, r_2, \dots, r_n\}$ of algebraic numbers such that $\prod_{i=1}^{n}(X - r_i)$ approximates $p(X)$. That is, μ provides a coherent method of approximately factoring $p(X)$ into linear factors. In practice, that's what you do: For any $\varepsilon > 0$, you calculate a factorization of $p(X)$ to within ε. The issue of coming up with a specific root, in the sense that you can identify (in advance) the approximation to it in any of these approximate factorizations, never arises. Thus the

construction of an element of $\widehat{M}_n(\mathbf{A})$ from $p(X)$ corresponds more closely to our numerical practice of the fundamental theorem of algebra than does constructing roots of $p(X)$.

Even in the presence of countable choice, we can't quite identify $\widehat{M}_n(\mathbf{C})$ with the space $M_n(\mathbf{C})$ of multisets of complex numbers. Recall the multisets of real numbers $\{a, b\}$ and $\{a \vee b, a \wedge b\}$ (where a and b are very close). We can't show that these multisets are equal, in the sense we can say that $a = a \vee b$ or $a = a \wedge b$, but the distance between them as elements of $\widehat{M}_2(\mathbf{C})$ is equal to zero, so they represent the same element of $\widehat{M}_2(\mathbf{C})$. They are each a multiset of roots of the polynomial $X^2 - (a + b)X + ab$. So even if we buy countable choice, there is a virtue to considering the set of roots of a polynomial to be an element of $\widehat{M}_n(\mathbf{C})$ — otherwise it is not uniquely defined.

As a variation on this theme of completing a space of finite sets, observe that a compact subset of a complete metric space X may be identified with an element of the completion $\widehat{F}$ of the set of finite subsets of X under the Hausdorff metric. Clearly any compact subset is arbitrarily close to a finite subset, and if the distance between two compact subsets is zero, then they are equal. However, countable choice is needed to pass from an element of $\widehat{F}$ to a compact subset of X. Indeed the element of $\widehat{M}_2(\mathbf{C})$ corresponding to the polynomial $X^2 - a$ also defines an element of the completion of $M_2(\mathbf{C})$ in the Hausdorff metric.

These compact sets that we get from $\widehat{F}$ can be approximated by finite sets, but not necessarily from within. This is similar to what happens classically in a space that is not complete — we can approximate things that are not in the space. The subset S of X corresponding to μ in $\widehat{F}$ consists of those points in X that are close to any finite subset that is close to μ in $\widehat{F}$. The problem is that a finite subset may be close to μ in $\widehat{F}$ but not (constructively) close to S in the Hausdorff metric.

Is it necessary that a compact subset be approximable *from within* by finite subsets? We could *define* a compact subset of X to be an element of $\widehat{F}$. If we do that, then the set of (complex) zeros of a polynomial is a compact subset, even though you may not be able to get hold of any of its elements. I claim that this is actually not contrary to our intuition regarding the set of zeros. We can, for any degree of approximation, find a set of approximate roots, even if we cannot specify a single root.

7. Homogenizing.

What about the restriction to monic polynomials? Bishop, after all, proved that any nonconstant polynomial with complex coefficients has a complex root, as did Brouwer [2]. In the general case, where the formal leading

coefficient might be zero, we really need to pass to the complex projective line (the Riemann sphere), and homogeneous polynomials in two variables. A polynomial whose formal leading coefficient is zero has a root at infinity. Indeed, van der Corput [4], following Brouwer [2], showed that you could factor an arbitrary nonzero polynomial $f(X) = a_0 + a_1X + \cdots + a_nX^n$ as

$$f(X) = a(X - r_1)\cdots(X - r_m)(1 - r_{m+1}X)\cdots(1 - r_nX)$$

where m is any integer between s and t such that a_s and a_t are nonzero.

Now the objects of study are nonzero polynomials of formal degree n, and we don't particularly want to distinguish between a polynomial and a nonzero multiple of that polynomial because the two polynomials have the same roots. So our polynomials live in $P^n(\mathbf{C})$, complex projective space of dimension n—the set of one-dimensional subspaces of $\mathbf{C}^{n+1}$. For any norm on $\mathbf{C}^{n+1}$, we define a metric on $P^n(\mathbf{C})$ by setting $d(U, V) = \inf \|u - v\|$, where u and v range over the elements of norm 1 of the one-dimensional subspaces U and V of $\mathbf{C}^{n+1}$. It is easy to verify that this metric is identical to the Hausdorff metric on the unit balls of the one-dimensional subspaces.

What about the roots? To accommodate points at infinity, we want to think of our polynomials as homogeneous polynomials of degree n in two variables. This is more elegant anyway because we can discard the idea of a formal degree. The roots will then be elements of $P^1(\mathbf{C})$, and correspond to homogeneous factors of degree 1.

What norm will we use? On $\mathbf{C}^2$ we will work with the ℓ_2-norm, $\sqrt{|a|^2+|b|^2}$. Then the distance between two unit vectors u and v in $\mathbf{C}^2$, viewed as elements of $P^1(\mathbf{C})$, is given by $\sqrt{2 - 2\,|\langle u, v\rangle|}$. Note that this norm on $\mathbf{C}^2$ is the same as that obtained by considering an element of $\mathbf{C}^2$ as a linear function $f(X, Y) = aX + bY$ and taking the supremum of $|f(u, v)|$ over all unit vectors (u, v) in $\mathbf{C}^2$. On $\mathbf{C}^{n+1}$, which we think of as the space of homogeneous polynomials in two variables of degree n, we will also use this functional norm, $\sup\{|f(u, v)| : |u|^2 + |v|^2 = 1\} = \sup\{|f(u, v)| : |u|^2 + |v|^2 \le 1\}$.

The functional norm on $\mathbf{C}^{n+1}$ is equivalent to the supremum norm used in the proof of Theorem 2. Indeed, suppose $f(X, Y) = \sum_{i=0}^{n} a_i X^i Y^{n-i}$. Let $N_0 = \sup_i |a_i|$ and $N_1 = \sup\{|f(u, v)| : |u|^2 + |v|^2 = 1\}$. Clearly $N_1 \le (n + 1)N_0$. To go the other way we note that, for $|z| \le 1$,

$$|f(z, 1)| = \sqrt{2}^n \left| f\left(\frac{z}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\right| \le \sqrt{2}^n N_1$$

while, by the Cauchy integral formula,

$$|a_k| = \left|\frac{1}{2\pi i}\oint \frac{f(z, 1)}{z^{k+1}} dz\right| \le \sup_{|z|=1} |f(z, 1)|$$

so $N_0 \le \sqrt{2}^n N_1$.

We have the commutative diagram

$$\begin{array}{ccc} (\mathbf{C}^2)^n & \to & \mathbf{C}^{n+1} \\ \downarrow & & \downarrow \\ P^1(\mathbf{C})^n & \to & P^n(\mathbf{C}) \end{array}$$

the top map being the n-fold product, where we think of $\mathbf{C}^2$ as consisting of homogeneous linear polynomials, and $\mathbf{C}^{n+1}$ as consisting of homogenous polynomials of degree n. The two side maps are the natural ones, and the bottom map is induced from the top one. If we restrict to algebraic numbers,

$$\begin{array}{ccc} (\mathbf{A}^2)^n & \to & \mathbf{A}^{n+1} \\ \downarrow & & \downarrow \\ P^1(\mathbf{A})^n & \to & P^n(\mathbf{A}) \end{array}$$

then we know that the top map is onto.

Now consider the left column of this diagram to be symmetrized—that is, the objects there are n-multisets of elements of $\mathbf{A}^2$ and $P^1(\mathbf{A})$. Because the top map is multiplication, this makes sense. Our task is to show that the bottom map is uniformly bicontinuous.

If $\alpha : \mathbf{A}^2 \to \mathbf{A}^2$ is a unitary transformation, then α induces isometries at all four corners of the diagram (we put the functional norm on A^{n+1}). We will use Theorem 2, together with this ability to move the point at infinity of $P^1(\mathbf{A})$, to show uniform bicontinuity of the bottom map.

Consider the map from $\mathbf{C}$ to $P^1(\mathbf{C})$ given by taking r to $(1, -r)$. This can be thought of as taking the complex number r to the homogeneous polynomial $X - rY$ whose corresponding inhomogeneous polynomial $X - r$ has r as a root. This map takes the closed ball of radius R around 0 in $\mathbf{C}$ uniformly bicontinuously onto the complement of the open ball of radius $\sqrt{2 - \dfrac{2R}{1+R^2}}$ around $\infty = (0, 1)$ in $P^1(\mathbf{C})$.

Now we have a bigger commutative diagram

$$\begin{array}{ccc} \mathbf{A}^n & \longleftrightarrow & \mathbf{A}^n \\ \downarrow & & \downarrow \\ (\mathbf{A}^2)^n & \to & \mathbf{A}^{n+1} \\ \downarrow & & \downarrow \\ P^1(\mathbf{A})^n & \to & P^n(\mathbf{A}) \end{array}$$

where the first column is symmetrized over n, that is, these are actually n-multisets of elements of $\mathbf{A}$, $\mathbf{A}^2$ and $P^1(\mathbf{A})$.

The top map takes $(r_1, \ldots, r_n)$ to the nonleading coefficients of the monic polynomial $\prod_{i=1}^n (X - r_i)$. This is the correspondence between $M_n(\mathbf{A})$ and $\pi_n(\mathbf{A})$ that we have seen is uniformly bicontinuous on bounded subsets (Theorem 2). The map down from $\mathbf{A}^n$ to $(\mathbf{A}^2)^n$ is induced by the map from $\mathbf{A}$ to $\mathbf{A}^2$ that takes r to $(1, -r)$. The map down from $\mathbf{A}^n$ to $\mathbf{A}^{n+1}$ takes $(c_0, c_1, \ldots, c_{n-1})$ to $(c_0, c_1, \ldots, c_{n-1}, 1)$.

So we know that the bottom map is uniformly bicontinuous if we stay bounded away from the point at infinity. The idea is to act on $P^1(\mathbf{A})$, with unitary transformations of $\mathbf{C}^2$, so that any n-multiset of roots can be uniformly bounded away from the point at infinity. As the (algebraic) unitary group acts transitively on $P^1(\mathbf{A})$, it suffices to find an element in $P^1(\mathbf{A})$ that is far away from a given n-multiset.

Lemma 5. *Let $z_1, \dots, z_n$ be elements of $\mathbf{A}^2$ of norm 1. Then there exists an element of $\mathbf{A}^2$ of norm 1 whose distance to $\{z_1, \dots, z_n\}$ in $P^1(\mathbf{A})$ is at least $1/2n$.*

Proof. This follows from the fact that $P^1(\mathbf{C})$ is connected, totally bounded, and has diameter $\sqrt{2}$, and that $P^1(\mathbf{A})$ is dense in $P^1(\mathbf{C})$. If F is a finite δ-approximation to $P^1(\mathbf{A})$, then either there exists an element of F that is at least $1/2n$ from each z_i, in which case we are done, or every element of F is within $\frac{1.1}{2n}$ of some z_i, so the z_i form a $\left(\frac{1.1}{2n} + \delta\right)$-approximation to P^1. As P^1 is connected, and has diameter $\sqrt{2}$, we can find a finite sequence of elements so that the distance between any two adjacent elements is less than δ, and the distance between the first and last is $\sqrt{2}$. We get a corresponding sequence of approximating z_i's so that the distance between adjacent elements is less than $\frac{1.1}{n} + 3\delta$. By throwing out cycles, we may assume that the i's are all distinct. But this would say that $\sqrt{2} < 1.1 + 3n\delta$, and we can choose δ so that does not happen. □

So, given elements $z_1, \dots, z_n$ of $\mathbf{A}^2$, each of norm 1, we can find a unitary transformation α of $\mathbf{A}^2$ so that $\alpha z_1, \dots, \alpha z_n$ are bounded away by $1/2n$ from the point at infinity in $P^1(\mathbf{A})$.

Here is the full-blown fundamental theorem of algebra.

Theorem 6. *Let $P^j(\mathbf{C})$ denote the set of nonzero homogeneous polynomials of degree j in two variables over the complex numbers, viewed as a projective space of dimension j. The map $\widehat{M}_n(P^1(\mathbf{C})) \to P^n(\mathbf{C})$ from the set of limit n-multisets of $P^1(\mathbf{C})$ to $P^n(\mathbf{C})$ that is induced by taking $(r_1X + s_1Y, r_2X + s_2Y, \dots, r_nX + s_nY)$ to $\prod_{i=1}^n (s_iX + r_iY)$, is a uniformly bicontinuous one-to-one correspondence.*

Proof. It suffices to show that the map $M_n(P^1(\mathbf{A})) \to P^n(\mathbf{A})$ is a uniformly bicontinuous one-to-one correspondence. Given the fundamental theorem of algebra for algebraic numbers, the only problem is the uniform bicontinuity. Suppose $(r_1X + s_1Y, r_2X + s_2Y, \dots, r_nX + s_nY)$ represents an element of $M_n(P^1(\mathbf{A}))$. As any unitary transformation of $\mathbf{A}^2$ induces an isometry on $P^j(\mathbf{A})$, under the functional norm, we may assume that $r_iX + s_iY \in P^1(\mathbf{A})$

is bounded away by $1/2n$ from Y, the point at infinity. In the diagram

$$\begin{array}{ccc} M_n(\mathbf{A}) & \longleftrightarrow & \mathbf{A}^n \\ \downarrow & & \downarrow \\ M_n(\mathbf{A}^2) & \to & \mathbf{A}^{n+1} \\ \downarrow & & \downarrow \\ M_n(P^1(\mathbf{A})) & \to & P^n(\mathbf{A}) \end{array}$$

the top map is uniformly bicontinuous on bounded subsets by Theorem 2. The image in $P^1(\mathbf{A})$ of the ball $\{z \in \mathbf{A} : |z| \leq R\}$ contains all those points of $P^1(A)$ that are bounded away from the point at infinity by $\sqrt{2 - \dfrac{2R}{1+R^2}}$, so for suitable R contains all those points bounded away from the point at infinity by $1/4n$. Thus we can find a modulus of bicontinuity for the bottom map that is independent of the chosen point. □

References

[1] E. Bishop, *Foundations of constructive analysis*, McGraw-Hill, 1967.

[2] L.E.J. Brouwer, *Intuitionistische Ergänzung des Fundamentalsatzes der Algebra*, Proc. Acad. Amsterdam, **27** (1924), 631-634.

[3] L.E.J. Brouwer and B. de Loor, *Intuitionistischer Beweis des Fundamentalsatzes der Algebra*, Proc. Acad. Amsterdam, **27** (1924), 186-188.

[4] J.G. van der Corput, *On the fundamental theorem of algebra*, Indagationes Mathematicae, **8** (1946), 430-440.

[5] A. Joyal and G.E. Reyes, *Separably real closed local rings*, Journal of Pure and Applied Algebra, **43** (1986), 271-279.

[6] B.A. Kushner, *Lectures on constructive mathematical analysis*, AMS Translations of Mathematical Monographs, **60** (1984).

[7] Ray Mines, F. Richman and W. Ruitenburg, *A course in constructive algebra*, Springer-Verlag, 1988.

[8] G.H. Moore, *Zermelo's axiom of choice, its origins, development and influence*, Springer-Verlag, 1982.

[9] F. Richman, *Intuitionism as generalization*, Philosophia Mathematica, **5** (1990), 124-128.

[10] ———, *Interview with a constructive mathematician*, Modern Logic, **6** (1996), 247-271.

[11] W.B.G. Ruitenburg, *Constructing roots of polynomials over the complex numbers*, Computational aspects of Lie group representations and related topics (Amsterdam, 1990), 107-128; CWI Tract, **84**, Math. Centrum, Centrum Wisk. Inform., Amsterdam, 1991.

[12] G. Stolzenberg, *Sets as limits*, preprint, 1988.

[13] A.S. Troelstra, *Intuitionistic extensions of the reals*, Nieuw Archief voor Wiskunde, **28** (1980), 63-113.

[14] A.S. Troelstra and Dirk van Dalen, *Constructivism in mathematics: An introduction*, North-Holland, 1988.

Received January 14, 1998 and revised October 28, 1998.

Florida Atlantic University
Boca Raton, FL 33431
E-mail address: richman@fau.edu