نظرية التقريب

المرحلة الرابعة
قسم الرياضيات
كلية العلوم للبنات

Introduction

In 1853, the great Russian mathematician, P. L. Chebyshev (Čebyšev), while working on a problem of *linkages*, devices which translate the linear motion of a steam engine into the circular motion of a wheel, considered the following problem:

> Given a continuous function f defined on a closed interval $[\,a, b\,]$ and a positive integer n, can we "represent" f by a polynomial $p(x) = \sum_{k=0}^{n} a_k x^k$, of degree at most n, in such a way that the maximum error at any point x in $[\,a, b\,]$ is controlled? In particular, is it possible to construct p so that the error $\max_{a \le x \le b} |f(x) - p(x)|$ is minimized?

This problem raises several questions, the first of which Chebyshev himself ignored:

– Why should such a polynomial even *exist*?

– If it does, can we hope to *construct* it?

– If it exists, is it also *unique*?

– What happens if we change the measure of error to, say, $\int_a^b |f(x) - p(x)|^2 \, dx$?

Def. 1.1.: If there is $a^* \in A$ such that $d(a^*, f) \leq d(a, f)$, $\forall\ a \in A$, then we say that a^* is a best approximation from A to *f*.

Theorem 1.2.: If A is a compact set in a metric space X, then for every $f \in X$, there exists an element $a^* \in A$ such that $d(a^*, f) \leq d(a, f)$, $\forall\ a \in A$, a^* is a best approximation from A to *f*.

Proof:
Let $d^* = \inf \{d(a, f) : a \in A\}$, if there exists $a^* \in A$ s.t. $d^* = d(a^*, f)$ then there is nothing to prove.
Otherwise, there is a sequence $\{a_i, i=1,2,..\}$ of points in A which gives the limit :

$$\lim_{i \to \infty} d(a_i, f) = d^*$$

By completeness, the sequence $\{a_i\}$ has at least one limit point in A (say a^*) such that: $\forall \varepsilon > 0, \exists k \in N$ s.t. $d(a_k, f) < d^* + \frac{1}{2}\varepsilon$

And $d(a_k, a^*) < \frac{1}{2}\varepsilon \quad (a_k \to a^*)$

Then $d(a^*, f) \le d(a^*, a_k) + d(a_k, f)$

$$< d^* + \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon$$

$$< d^* + \varepsilon$$

Since ε is an arbitrary real valued $\to$ $d(a^*, f) < d(a, f)$

Then a^* is a best approximation element from A to f.

Introduction

In 1853, the great Russian mathematician, P. L. Chebyshev (Čebyšev), while working on a problem of *linkages*, devices which translate the linear motion of a steam engine into the circular motion of a wheel, considered the following problem:

> Given a continuous function f defined on a closed interval $[\,a, b\,]$ and a positive integer n, can we "represent" f by a polynomial $p(x) = \sum_{k=0}^{n} a_k x^k$, of degree at most n, in such a way that the maximum error at any point x in $[\,a, b\,]$ is controlled? In particular, is it possible to construct p so that the error $\max_{a \le x \le b} |f(x) - p(x)|$ is minimized?

This problem raises several questions, the first of which Chebyshev himself ignored:

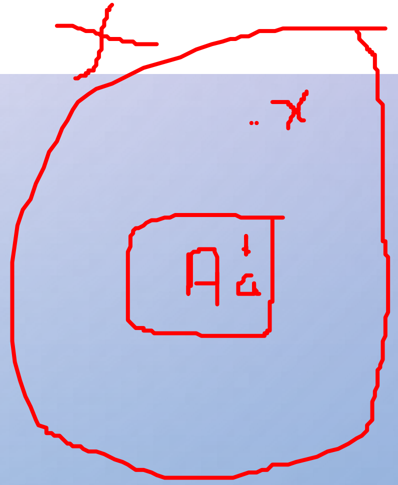
– Why should such a polynomial even *exist*?

– If it does, can we hope to *construct* it?

– If it exists, is it also *unique*?

– What happens if we change the measure of error to, say, $\int_a^b |f(x) - p(x)|^2\,dx$?

We have a metric space X and we want to approximate a given element $x \in X$ by an element a of some subset A of X. The elements of A are "nice" or "tractable" and we want to make the distance between x and a as small as possible: we call this "to make a good approximation of x by elements of A."



To any element x of X and any subset A of X we associate the distance $d(x, A)$ from x to A, which by definition is

$$d(x, A) = \inf_{a \in A} d(x, a), \qquad x \in X,\ A \subset X. \tag{1.4}$$

Obviously we have $0 \leqslant d(x, A) \leqslant +\infty$ with equality at the second place if and only if A is empty and equality at the first place if and only if a belongs to $\overline{A}$, the closure of A. So the elements x such that $d(x, A) = 0$ are those which can be approximated arbitrarily well by nice elements. If $d(x, A) > 0$ there is a certain unavoidable error.

A very common situation is that we have an increasing sequence (A_m) of sets whose union is dense in X, so that $d(x, A_m) \to 0$ as $m \to \infty$ for every $x \in X$. Then an interesting question is how fast the convergence is and how the rate of convergence depends on properties of the element x.

It may or may not happen that the infimum in (1.4) is a minimum. In other words, it may happen that there exists an element a, called *a best approximant*, such that

$$d(x, a) = d(x, A),$$

but it may also be the case that

$$d(x, a) > d(x, A) \text{ for all } a \in A.$$

In the latter case we are interested in constructing a sequence (a_j) of elements of A such that $d(x, a_j) \to d(x, A)$ as $j \to \infty$. We call such a sequence *an approximating sequence*. In the first case we may ask if there is a unique best approximant: the set

$$\{a \in A; d(x,a) = d(x,A)\},$$

may be empty, have exactly one element, or may have more than one element.

Definition 1.1.: If there is $a^* \in A$ such that $d(a^*, f) \leq d(a, f)$, $\forall\ a \in A$, then we say that a^* is a best approximation from A to f.

Theorem 1.2.: If A is a compact set in a metric space X, then for every $f \in X$, there exists an element $a^* \in A$ such that $d(a^*, f) \le d(a, f)$, $\forall\, a \in A$, a^* is a best approximation from A to f.

Proof:

Let $d^* = \inf \{d(a, f) : a \in A\}$, if there exists $a^* \in A$ s.t. $d^* = d(a^*, f)$ then there is nothing to prove.

Otherwise, there is a sequence $\{a_i, i=1,2,..\}$ of points in A which gives the limit :

$$\lim_{i \to \infty} d(a_i, f) = d^*$$

By completeness, the sequence $\{a_i\}$ has at least one limit point in A (say a^*) such that:

$$\forall \varepsilon > 0, \exists k \in N \ \text{ s.t. } \ d(a_k, f) < d^* + \frac{1}{2}\varepsilon$$

And $d(a_k, a^*) < \frac{1}{2}\varepsilon \quad (a_k \to a^*)$

Then $d(a^*, f) \le d(a^*, a_k) + d(a_k, f)$

$$< d^* + \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon$$

$$< d^* + \varepsilon$$

Since ε is an arbitrary real valued $\to$ $d(a^*, f) < d(a, f)$

Then a^* is a best approximation element from A to f.

Best Approximations in Normed Spaces

Chebyshev's problem is perhaps best understood by rephrasing it in modern terms. What we have here is a problem of *best approximation* in a *normed linear space*. Recall that a *norm* on a (real) *vector space* X is a nonnegative function on X satisfying

$\|x\| \geq 0$, and $\|x\| = 0$ if and only if $x = 0$,

$\|\alpha x\| = |\alpha|\|x\|$ for any $x \in X$ and $\alpha \in \mathbb{R}$,

$\|x + y\| \leq \|x\| + \|y\|$ for any $x, y \in X$.

Any norm on X induces a *metric* or distance function by setting $\text{dist}(x, y) = \|x - y\|$. The abstract version of our problem(s) can now be restated:

Lemma 1.3. *Let V be a finite-dimensional vector space. Then, all norms on V are equivalent. That is, if $\|\cdot\|$ and $|||\cdot|||$ are norms on V, then there exist constants $0 < A, B < \infty$ such that*

$$A\,\|x\| \leq |||x||| \leq B\,\|x\|$$

for all vectors $x \in V$.

Corollary 1.4. *Every finite-dimensional normed space is complete (that is, every Cauchy seqeuence converges). In particular, if Y is a finite-dimensional subspace of a normed linear space X, then Y is a closed subset of X.*

Corollary 1.5. *Let Y be a finite-dimensional normed space, let $x \in Y$, and let $M > 0$. Then, any closed ball $\{ y \in Y : \|x - y\| \le M \}$ is compact.*

Theorem 1.6. *Let Y be a finite-dimensional subspace of a normed linear space X, and let $x \in X$. Then, there exists a (not necessarily unique) vector $y^* \in Y$ such that*

$$\|x - y^*\| = \min_{y \in Y} \|x - y\|$$

for all $y \in Y$. That is, there is a best approximation to x by elements from Y.

Proof. First notice that because $0 \in Y$, we know that any nearest point y^* will satisfy $\|x - y^*\| \le \|x\| = \|x - 0\|$. Thus, it suffices to look for y^* in the *compact* set

$$K = \{ y \in Y : \|x - y\| \le \|x\| \}.$$

To finish the proof, we need only note that the function $f(y) = \|x - y\|$ is *continuous*:

$$|f(y) - f(z)| = \Big| \|x - y\| - \|x - z\| \Big| \le \|y - z\|,$$

and hence attains a minimum value at some point $y^* \in K$. □

Corollary 1.7. *For each $f \in C[a,b]$ and each positive integer n, there is a (not necessarily unique) polynomial $p_n^* \in \mathcal{P}_n$ such that*

$$\|f - p_n^*\| = \min_{p \in \mathcal{P}_n} \|f - p\|.$$

Lemma 1.8 *Let Y be a finite-dimensional subspace of a normed linear space X, and suppose that each $x \in X$ has a unique nearest point $y_x \in Y$. Then the nearest point map $x \mapsto y_x$ is continuous.*

Proof. Let's write $P(x) = y_x$ for the nearest point map, and let's suppose that $x_n \to x$ in X. We want to show that $P(x_n) \to P(x)$, and for this it's enough to show that there is a subsequence of $(P(x_n))$ that converges to $P(x)$. (Why?)

Because the sequence (x_n) is bounded in X, say $\|x_n\| \le M$ for all n, we have

$$\|P(x_n)\| \le \|P(x_n) - x_n\| + \|x_n\| \le 2\|x_n\| \le 2M.$$

Thus, $(P(x_n))$ is a bounded sequence in Y, a finite-dimensional space. As such, by passing to a subsequence, we may suppose that $(P(x_n))$ converges to some element $P_0 \in Y$. (How?) Now we need to show that $P_0 = P(x)$. But

$$\|P(x_n) - x_n\| \le \|P(x) - x_n\|$$

for any n. (Why?) Hence, letting $n \to \infty$, we get

$$\|P_0 - x\| \le \|P(x) - x\|.$$

Because nearest points in Y are unique, we must have $P_0 = P(x)$. □

Theorem 1 .9 *Let Y be a subspace of a normed linear space X, and let $x \in X$. The set Y_x, consisting of all best approximations to x out of Y, is a bounded* convex *set.*

Proof. As we've seen, the set Y_x is a subset of the ball $\{ y \in X : \|x-y\| \leq \|x\| \}$ and, as such, is bounded. (More generally, the set Y_x is a subset of the sphere $\{ y \in X : \|x-y\| = d \}$, where $d = \text{dist}(x, Y) = \inf_{y \in Y} \|x - y\|$.)

Next recall that a subset K of a vector space V is said to be *convex* if K contains the line segment joining any pair of its points. Specifically, K is convex if

$$x, y \in K,\ 0 \leq \lambda \leq 1 \implies \lambda x + (1-\lambda) y \in K.$$

Thus, given $y_1, y_2 \in Y_x$ and $0 \leq \lambda \leq 1$, we want to show that the vector $y^* = \lambda y_1 + (1-\lambda) y_2 \in Y_x$. But $y_1, y_2 \in Y_x$ means that

$$\|x - y_1\| = \|x - y_2\| = \min_{y \in Y} \|x - y\|.$$

Hence,

$$\begin{aligned} \|x - y^*\| &= \|x - (\lambda y_1 + (1-\lambda) y_2)\| \\ &= \|\lambda(x - y_1) + (1-\lambda)(x - y_2)\| \\ &\leq \lambda \|x - y_1\| + (1-\lambda)\|x - y_2\| \\ &= \min_{y \in Y} \|x - y\|. \end{aligned}$$

Consequently, $\|x - y^*\| = \min_{y \in Y} \|x - y\|$; that is, $y^* \in Y_x$. □

A norm $\|\cdot\|$ on a vector space X is said to be *strictly convex* if, for any pair of points $x \neq y \in X$ with $\|x\| = r = \|y\|$, we always have $\|\lambda x + (1-\lambda)y\| < r$ for all $0 < \lambda < 1$. That is, the open line segment between any pair of points on the sphere of radius r lies entirely within the open ball of radius r; in other words, only the endpoints of the line segment can hit the sphere. For simplicity, we often say that the space X is strictly convex, with the understanding that we're actually referring to a property of the norm in X. In any such space, we get an immediate corollary to our last result:

Corollary 1 .10 *If X has a strictly convex norm, then, for any subspace Y of X and any point $x \in X$, there can be at most one best approximation to x out of Y. That is, Y_x is either empty or consists of a single point.*

Lemma 1. .11 *A normed space X has a strictly convex norm if and only if the triangle inequality is strict on nonparallel vectors; that is, if and only if*

$$x \neq \alpha y, \;\; y \neq \alpha x, \;\; \text{all } \alpha \in \mathbb{R} \implies \|x + y\| < \|x\| + \|y\|.$$

Proof. First suppose that X is strictly convex, and let x and y be nonparallel vectors in X. Then, in particular, the vectors $x/\|x\|$ and $y/\|y\|$ must be different. (Why?) Hence,

$$\left\| \left(\frac{\|x\|}{\|x\| + \|y\|}\right) \frac{x}{\|x\|} + \left(\frac{\|y\|}{\|x\| + \|y\|}\right) \frac{y}{\|y\|} \right\| < 1.$$

That is, $\|x + y\| < \|x\| + \|y\|$.

Next suppose that the triangle inequality is strict on nonparallel vectors, and let $x \neq y \in X$ with $\|x\| = r = \|y\|$. If x and y are parallel, then we must have $y = -x$. (Why?) In this case,

$$\|\lambda x + (1-\lambda)\, y\| = |2\lambda - 1|\, \|x\| < r$$

because $-1 < 2\lambda - 1 < 1$ whenever $0 < \lambda < 1$. Otherwise, x and y are nonparallel. Thus, for any $0 < \lambda < 1$, the vectors λx and $(1-\lambda)\, y$ are likewise nonparallel and we have

$$\|\lambda x + (1-\lambda)\, y\| < \lambda\|x\| + (1-\lambda)\|y\| = r. \qquad \square$$

Examples 1 .12

1. The usual norm on $C[a,b]$ is *not* strictly convex (and so the problem of uniqueness of best approximations is all the more interesting to tackle). For example, if $f(x) = x$ and $g(x) = x^2$ in $C[0,1]$, then $f \neq g$ and $\|f\| = 1 = \|g\|$, while $\|f+g\| = 2$. (Why?)

2. The usual norm on $\mathbb{R}^n$ is strictly convex, as is any one of the norms $\|\cdot\|_p$ for $1 < p < \infty$. (See Problem 10.) The norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$, on the other hand, are *not* strictly convex. (Why?)