Chebyshev Criterion

written by Oleg Ivrii

Given two continuous functions f, g on an interval $[a, b]$, we will measure their distance by taking $||f - g|| = \max_{x \in [a,b]} |f(x) - g(x)|$. We are interested in the following problem:

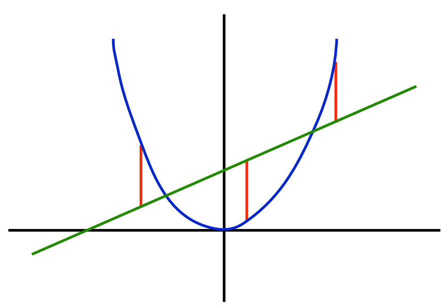
Given a continuous function $f(x)$ on interval $[a, b]$, find the best approximating polynomial $P(x)$, i.e one which minimizes $||f - P||$. More specifically, we are interested if the best approximating polynomial exists, if it is unique and what distinguishing properties it has.

A famous theorem of Weierstrass tells us that if f is not already a polynomial, f can be approximated arbitrarily well in uniform norm, so we must rephrase the question: *Find the best approximating polynomial of degree at most n.*

Chebyshev discovered that such a best approximating polynomial exists, is unique and moreover satisfies the following property: *there exist ripple points $x_1, x_2, \dots, x_{n+2} \in [a, b]$ such that*

$$||f - P|| = |f(x_1) - P(x_1)| = |f(x_2) - P(x_2)| = \cdots = |f(x_{n+2} - P(x_{n+2})|$$

and furthermore $(f - P)(x_1) = -(f - P)(x_2) = (f - P)(x_3) = \cdots = \pm(f - P)(x_{n+2})$.



For example, if a line approximates a parabola, it should have three ripple points.

It also turns out that Chebyshev's criterion is sufficient, that is any polynomial which does satisfy the Chebyshev criterion is necessarily best approximating.

Notation: we will denote the continuous functions on $[a, b]$ by $\mathscr{C}([a, b])$, polynomials by $\mathscr{P}([a, b])$ and polynomials of degree at most n by $\mathscr{P}^n([a, b])$.

Existence of the best approximating polynomial

Polynomials $\mathscr{P}^n([a,b])$ can be thought of as continuous functions of the coefficients. Consider the map $\mathscr{P} : \mathbb{R}^{n+1} \to \mathscr{C}([a,b])$ which given a tuple $(a_0, a_1, a_2, \dots a_n)$ constructs the polynomial $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$. We just get a mapping $\mathscr{P} : \mathbb{R}^{n+1} \to \mathscr{C}([a,b])$. This mapping is *continuous.* This means that if the coefficients are changed slightly, then the polynomial won't change much in the uniform norm on $[a,b]$.

We want to show that we can restrict ourselves to polynomials with small coefficients, that is polynomials with large coefficients are not good candidates in approximating f. More specifically, we show that *if P has a coefficient of absolute value greater or equal to M then $||P|| \geq kM$ for some constant k.* In particular, this means that if $kM > 2||f||$, then P provides worse approximation to f than the trivial polynomial which is identically zero, i.e $||f - P|| \geq ||P|| - ||f|| \geq ||f|| = ||f - 0||$. Having shown this claim, existence of the best approximating polynomial would follow from compactness of $[-M, M]^{n+1}$.

The proof of the claim itself is deceptively simple: consider the boundary of the cube $Q = [-1,1]^{n+1} \subset \mathbb{R}^{n+1}$. It represents polynomials whose largest coefficient is 1. As none of these polynomials are 0 identically, by compactness

$$\min_{\{a_i\} \in \partial Q} ||\mathscr{P}(a_0, a_1, \dots, a_n)|| = k > 0.$$

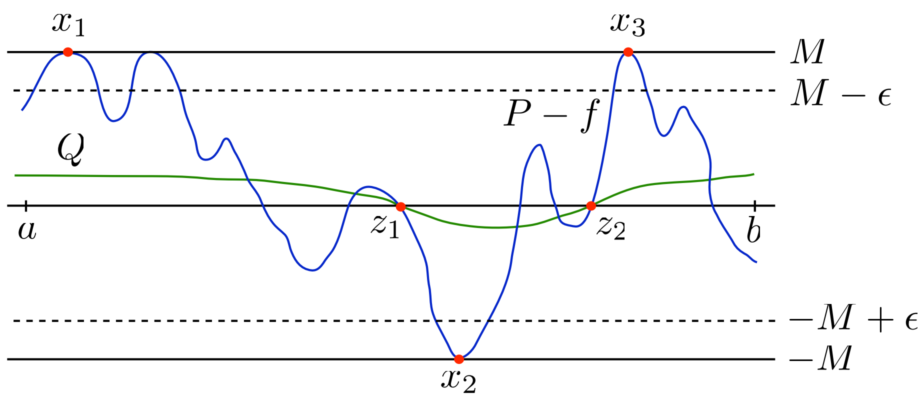
It remains to notice that $\mathscr{P}$ is *linear*, that is $\mathscr{P}(\lambda a_0, \lambda a_1, \dots \lambda a_n) = \lambda P(a_0, a_1, \dots, a_n)$: if the largest coefficient of a polynomial P is M, $||P|| \geq kM$.

Sufficiency of the Chebyshev Criterion

Now, we prove a polynomial P satisfying the Chebyshev criterion is necessarily best approximating. Suppose polynomial Q was such that $||f - Q|| < ||f - P||$. Consider the polynomial $P - Q$. Suppose that at x_{odd}, $P > f$ while at x_{even}, $P < f$. Then $P - Q$ is strictly positive at x_{odd} and negative at x_{even}. This shows that that $P - Q$ has zeros between all ripple points, so has at least $n+1$ zeros. But it is a polynomial of degree at most n, so it must be identically 0. Thus $P = Q$ identically.

Necessity of the Chebyshev Criterion

We must understand what it means for a polynomial P not to satisfy the Chebyshev criterion. For this purpose we construct a special partition $a, z_1, z_2, \ldots, z_k, b$ of the interval $[a, b]$. On each interval $[a, z_1], [z_1, z_2], \ldots, [z_{k-1}, z_k], [z_k, b]$ will be a point x_j for which $|P(x_j) - f(x_j)| = M$ where $M = ||P - f||$. If on an interval $[z_j, z_{j+1}]$ there is a point x_j for which $P(x_j) - f(x_j) = M$, we call this interval *mostly positive*. Otherwise, we call it *mostly negative*. We will require that on mostly positive intervals that $P - f$ would not attain the value $-M$, so it will be at least $-M + \epsilon$ for some ϵ. Similarly, on mostly negative intervals, we would require that $P - f$ does not attain the value M and so it will be at most $M - \epsilon$. We would also like mostly positive and mostly negative intervals to alternate.



We first choose the points x_j. Starting at a, we scan right for the first point x_1 where $|P(x_1) - f(x_1)| = M$. Next we look for the first point x_2 where $|P(x_2) - f(x_2)|$ also equals M but of opposite sign than $P(x_1) - f(x_1)$. We keep scanning for alternating minima and maxima until the scanner reaches b (the interval has been exhausted). Now on each interval $[x_j, x_{j+1}]$, let z_j be the *last* point for which $P(z) - f(z) = 0$.

The fact that P does not satisfy the Chebyshev conditions means that the number of points x_j is at most $n + 1$, i.e the number of z_j is at most n. Consider the polynomial $(x - z_1)(x - z_2) \ldots (x - z_k)$. By multiplying it by a small constant, we can insist that its norm is less than ϵ. Also, multiply it by ± 1 so that it "matches up" with $P - f$, i.e we want it to be positive whenever $P - f$ is mostly positive and negative when $P - f$ is mostly negative. Call the resulting polynomial Q. By construction, the degree of Q is at most n, so $P - Q$ is a better approximation to f than P.

Uniqueness of the best approximating polynomial

Suppose we had two best approximating polynomials P, Q. Let $M = ||f - P|| = ||f - Q||$. It is clear that any convex combination of P and Q say $\frac{P+Q}{2}$ is also best approximating. But $\frac{P+Q}{2}$ satisfies the Chebyshev criterion, so there must exist $n+2$ points x_i where $|\frac{P(x_i)+Q(x_i)}{2} - f(x_i)| = M$. But this is possible only if $P(x_i) = Q(x_i)$. This would mean that P and Q coincide at $n+2$ points, but seeing that they are polynomials of degree at most n, they would have to coincide everywhere.

Chebyshev Polynomial

The Chebyshev polynomial $T_n(x)$ is defined to be $\cos(n \arccos x)$. It is easily seen that $T_0(x) = 1, T_1(x) = x, T_2(x) = 2x - 1, T_3(x) = 4x^2 - 3x + 1$ and more generally from the recurrence relation $T_{n+2}(x) = 2xT_n(x) - T_{n-1}(x)$, $T_n(x)$ is a polynomial of degree n and leading coefficient 2^{n-1} for $n \geq 1$.

This section is devoted to showing how the Chebyshev polynomial $T_n(x) = \cos(\arccos nx)$ shows up in the problem of *finding the polynomial* $P(x)$ *of degree at most* $n-1$ *which best approximates* x^n *uniformly on the interval* $[-1, 1]$.

We draw information about $P(x)$ from the Chebyshev criterion. We are guaranteed the existence of points $x_1 < x_2 < \cdots < x_{n+2}$ for which $|x^n - f(x)| = ||x^n - f(x)|| = M$ with $|x_j^n - f(x_j)|$ alternating in sign. We will write $F(x) = x^n - P(x)$.

The polynomial $F(x)^2$ obtains the value M^2 at the points $x_1, x_2, \ldots, x_{n+1}$. For a second, suppose that $x_1 \neq -1$ and $x_{n+1} \neq 1$. Then all points $x_1, x_2, \ldots, x_{n+2}$ are local maxima forcing $F(x)^2$ to have degree at least $2n+2$. But clearly $F(x)^2$ has degree only $2n$. Thus $x_1 = -1, x_{n+1} = 1$ and $F(x)^2$ achieves the value M^2 with multiplicity one at x_1 and x_{n+1} and with multiplicity two at the other ripple points. As the leading coefficient of $F(x)$ is 1, this tells us that $F(x)^2 - M^2 = (x-1)(x+1)\prod_{j=1}^{n+1}(x - x_j)^2$.

We now consider the polynomial $F'(x)$. It has degree $n-2$ which must be accounted by the local maxima at $x_2, \ldots, x_{n+1}$. As the leading coefficient of $F'(x)$ is n, we see that $F'(x) = n\prod_{j=1}^{n+1}(x - x_j)$. Thus we come to the differential equation

$$n^2(M^2 - F(x)^2) = (1 - x^2)F'(x)^2.$$

To solve the differential equation, we first take square roots of both sides and integrate. This turns out to be quite delicate.

Taking Square Roots

Care must be taken as $F'(t)$ changes sign. If $t \in [-1,1]$, let $j(t)$ mean the index j for which the interval $[x_j, x_{j+1})$ contains t. Notice that if $n-j$ is even, while t goes from x_j to x_{j+1}, $F(t)$ increases from $-M$ to M and $F'(t) > 0$; however, if $n-j$ is odd, $F(t)$ would decrease from M to $-M$ and $F'(t) < 0$.

We thus obtain $n\sqrt{M^2 - F(t)^2} = \sqrt{t^2-1} \cdot (-1)^{n-j(t)} F'(t)$. Then

$$n \frac{dt}{\sqrt{t^2-1}} = (-1)^{n-j(t)} \frac{F'(t)dt}{\sqrt{M^2 - F(t)^2}}. \tag{1}$$

Integrating both sides

We integrate both sides from x to 1. Recall that $\arccos x = \int_x^1 \frac{dt}{\sqrt{1-t^2}}$ where $\arccos t$ is the inverse of $\cos : [0,\pi] \to [-1,1]$. Additionally, the full integral $\int_{-1}^1 \frac{dt}{\sqrt{1-t^2}} = \pi$. Also set $w(t) = \arccos(F(t)/M)$. Then $w'(t) = \frac{F'(t)}{\sqrt{M^2-F(t)^2}}$.

Claim: upon integration from x to 1, we obtain the formula

$$n \arccos x = (-1)^{n-j} \arccos(F(x)/M) + 2\pi k \tag{2}$$

where $j = j(x)$ and k is some integer to be determined.

Integration of the LHS of (1) is straightforward but integrating the RHS is tricky.

$$\int_x^1 (-1)^{n-j(t)} w'(t)dt = (-1)^{n-j} \int_x^{x_{j+1}} w'(t)dt + \sum_{s=j+1}^{n} \int_{x_s}^{x_{s+1}} (-1)^{n-s} w'(t)dt.$$

If $n-j$ is even, changing variables we see $\int_x^{x_{j(x)+1}} w'(t)dt = \int_{F(x)/M}^1 \frac{dt}{\sqrt{1-t^2}} = \arccos(F(x)/M)$. Alternatively, if $n-j$ is odd, $\int_x^{x_{j(x)+1}} w'(t)dt = \int_{F(x)/M}^{-1} \frac{dt}{\sqrt{1-t^2}} = \arccos(F(x)/M) - \pi$. Similar considerations tells us that every term in the sum equals $+\pi$ giving a total to $(n-j)\pi$. Thus (2) is satisfied with $k = \lfloor \frac{n-j+1}{2} \rfloor$.

Taking cosines of both sides of (2), both the $(-1)^{n-j}$ factor and $+2\pi k$ will disappear and we find that $T_n(x) = \cos(n \arccos x) = F(x)/M$, i.e. $P(x) = x^n - MT_n(x)$. As $P(x)$ has degree $n-1$, we must have $M = 1/2^{n-1}$ and so $P(x) = x^n - T_n(x)/2^{n-1}$.