Notes on how to prove
Chebyshev's equioscillation theorem

The main issue is to characterize the minimax approximation p of a continuous function $f\colon [a,b] \to \mathbb{R}$ over all polynomials p with $\deg p \le n$. Chebyshev showed that this could be achieved if $f-p$ had the following "equioscillation" property: there are $n+2$ points $x_0,\, x_1,\, x_2,\, \ldots,\, x_{n+1}$ with $a \le x_0 < x_1 < x_2 < \cdots < x_n < x_{n+1} \le b$ and

$$f(x_i) - p(x_i) \;\; = \;\; \sigma\,(-1)^i\, \|f-p\|_\infty\,, \qquad \sigma = \pm 1.$$

This is not too difficult once we learn how to take "derivatives" of $g \mapsto \|g\|_\infty := \max_{a\le x\le b} |g(x)|$. The trouble is that this is not a smooth functions, and does not have ordinary derivatives everywhere. It is, however, a convex function and so it has *directional* derivatives. This will not be proved here, but we will take it for a fact. (It is — or can — be proved in Optimization Techniques, 22M:174.)

Let $\phi(g) = \|g\|_\infty$. Now the direction derivative in the direction r where $r\colon [a,b] \to \mathbb{R}$ (continuous) is

$$\phi'(g;r) \;\; := \;\; \lim_{\epsilon \downarrow 0} \frac{\phi(g+\epsilon r) - \phi(g)}{\epsilon}.$$

Now $\|g+\epsilon r\|_\infty = |g(x_\epsilon) + \epsilon r(x_\epsilon)|$ for some $x_\epsilon \in [a,b]$. We will show that

$$\phi'(g;r) \;\; = \;\; \max_{x \in \mathcal{M}(g)} \text{sign}(g(x))\, r(x)$$

where $\mathcal{M}(g) = \{\, x \in [a,b] \mid |g(x)| = \|g\|_\infty \,\}$, at least assuming that g is not everywhere zero.

We want to compute the limit $\left(|g(x_\epsilon) + \epsilon\, r(x_\epsilon)| - \|g\|_\infty\right)/\epsilon$. Now

$$\begin{aligned} |g(x_\epsilon) + \epsilon\, r(x_\epsilon)| \;\; &= \;\; \text{sign}\left(g(x_\epsilon) + \epsilon\, r(x_\epsilon)\right)\, \left(g(x_\epsilon) + \epsilon\, r(x_\epsilon)\right) \\ &= \;\; \text{sign}\left(g(x_\epsilon)\right)\, \left(g(x_\epsilon) + \epsilon\, r(x_\epsilon)\right) \end{aligned}$$

since, $g(x_\epsilon)$ is bounded away from zero. To see this, note that

$$\begin{aligned} |g(x_\epsilon)| \;\; &\ge \;\; |g(x_\epsilon) + \epsilon\, r(x_\epsilon)| - \epsilon\, |r(x_\epsilon)| \\ &= \;\; \|g + \epsilon\, r\|_\infty - \epsilon\, |r(x_\epsilon)| \\ &\ge \;\; \|g\|_\infty - 2\epsilon\, \|r\|_\infty\,; \end{aligned}$$

provided $\|g\|_\infty > 3\epsilon \|r\|_\infty$ we have $|g(x_\epsilon)| > \epsilon |r(x_\epsilon)|$ and so the signs of $g(x_\epsilon)$ and $g(x_\epsilon) + \epsilon r(x_\epsilon)$ are the same. Also $|g(x_\epsilon)| \to \|g\|_\infty$ as $\epsilon \downarrow 0$.

Now $x_\epsilon \in [a, b]$ for all $\epsilon > 0$, so since $[a, b]$ is a compact set, there is a convergent subsequence (also denoted x_ϵ) which has a limit $\widehat{x}$. Since g is continuous, $|g(\widehat{x})| = \lim_{\epsilon\downarrow 0} |g(x_\epsilon)| = \|g\|_\infty$ (taking the limit in the subsequence). So $\widehat{x} \in \mathcal{M}(g)$.

Now,

$$\begin{aligned} \|g + \epsilon r\|_\infty &= |g(x_\epsilon) + \epsilon r(x_\epsilon)| \\ &= \textsf{sign}(g(x_\epsilon) + \epsilon r(x_\epsilon))\ (g(x_\epsilon) + \epsilon r(x_\epsilon)) \\ &= \textsf{sign}(g(x_\epsilon))\ (g(x_\epsilon) + \epsilon r(x_\epsilon)) \\ &= |g(x_\epsilon)| + \textsf{sign}(g(x_\epsilon))\, \epsilon\, r(x_\epsilon) \\ &\le \|g\|_\infty + \epsilon\, \textsf{sign}(g(x_\epsilon))\, r(x_\epsilon). \end{aligned}$$

Therefore, for $\epsilon > 0$,

$$\begin{aligned} \frac{\|g + \epsilon r\|_\infty - \|g\|_\infty}{\epsilon} &\le \frac{\|g\|_\infty + \epsilon\, \textsf{sign}(g(x_\epsilon))\, r(x_\epsilon) - \|g\|_\infty}{\epsilon} \\ &= \textsf{sign}(g(x_\epsilon))\, r(x_\epsilon) \to \textsf{sign}(g(\widehat{x}))\, r(\widehat{x}) \end{aligned}$$

as $\epsilon \downarrow 0$. Thus $\phi'(g; r) \le \textsf{sign}(g(\widehat{x}))\, r(\widehat{x}) \le \max_{x\in\mathcal{M}(g)} \textsf{sign}(g(x))\, r(x)$.

On the other hand, for $\epsilon > 0$ sufficiently small and $\widehat{x} \in \mathcal{M}(g)$,

$$\begin{aligned} \frac{\|g + \epsilon r\|_\infty - \|g\|_\infty}{\epsilon} &\ge \frac{|g(\widehat{x}) + \epsilon r(\widehat{x})| - |g(\widehat{x})|}{\epsilon} \\ &= \frac{\textsf{sign}(g(\widehat{x}))\ (g(\widehat{x}) + \epsilon r(\widehat{x})) - |g(\widehat{x})|}{\epsilon} \\ &= \frac{|g(\widehat{x})| + \epsilon\, \textsf{sign}(g(\widehat{x}))\, r(\widehat{x}) - |g(\widehat{x})|}{\epsilon} \\ &= \textsf{sign}(g(\widehat{x}))\, r(\widehat{x}). \end{aligned}$$

Since this is true for all $\widehat{x} \in \mathcal{M}(g)$,

$$\frac{\|g + \epsilon r\|_\infty - \|g\|_\infty}{\epsilon} \ge \max_{x\in\mathcal{M}(g)} \textsf{sign}(g(x))\, r(x).$$

Since we have inequalities going both ways,

$$\phi'(g; r) = \max_{x\in\mathcal{M}(g)} \textsf{sign}(g(x))\, r(x).$$

Finishing the proof

To finish the proof, we first note that there *is* an optimum polynomial p of degree $\leq n$: if $\|p\|_\infty \to \infty$ then $\|f - p\|_\infty \to \infty$ as well. To search for an optimum we need consider only $\|p\|_\infty \leq \|f\|_\infty$. This in turn bounds the coefficients of p. Then the set of possible p is a bounded, closed, finite-dimensional set, and so is compact. Thus a minimizer exists.

At such a minimizer we need $\phi'(f-p; r) \geq 0$ for *any* polynomial r of degree $\leq n$. That is, for any polynomial of degree $\leq n$ we must have

$$\max_{x \in \mathcal{M}(f-p)} \text{sign}(f(x) - p(x))\, r(x) \geq 0.$$

If the number of points in $\mathcal{M}(f - p)$ is less than $n + 2$, then we can simply make $r(x)$ the polynomial interpolant of $r(x_i) = -\text{sign}(f(x_i)-p(x_i))$ where $\mathcal{M}(f - p) = \{x_0, x_1, \ldots, x_k\}$. Then this condition fails. So we must have at least $n + 2$ points in $\mathcal{M}(f - p)$.

To show the oscillation property, suppose that $f(x) - p(x)$ changes sign only k times on $\mathcal{M}(f - p)$ with $k < n + 1$. Then pick the points z_i to lie between the points where $f(x) - p(x)$ changes sign on $\mathcal{M}(f - p)$, and set $r(x) = \pm(x - z_1)(x - z_2)\cdots(x - z_k)$ so that $\deg r = k \leq n$. Choosing the sign appropriately we can make $\text{sign}(f(x) - p(x))\, r(x) < 0$ for every $x \in \mathcal{M}(f-p)$. Thus $f(x)-p(x)$ changes sign at least $n+1$ times on $\mathcal{M}(f - p)$. Thus there are at least $n + 2$ points x_i ($x_{i+1} > x_i$) in $\mathcal{M}(f - p)$ where $f(x_i) - p(x_i)$ alternate in sign as i increases; this is precisely Chebyshev's equioscillation theorem.