NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

For instructions about submitting Notes for publication in this department see the inside front cover.

A PROBABILISTIC PROOF OF THE WEIERSTRASS APPROXIMATION THEOREM

KENNETH M. LEVASSEUR
Department of Mathematics, University of Lowell, Lowell, MA 01854

Among the proofs of the Weierstrass Approximation Theorem, Bernstein's proof is surely the most common. It is particularly appealing because it is a constructive proof. If f is a continuous real-valued function of the interval $[0,1]$, the sequence of Bernstein polynomials

$$B_n(f;x) = \sum_{j=0}^{n} f\left(\frac{j}{n}\right)\binom{n}{j}x^j(1-x)^{n-j}, \quad n = 0,1,2,\ldots$$

converges uniformly to f. This fact establishes

THE WEIERSTRASS APPROXIMATION THEOREM. *If f is a continuous real-valued function on the interval $[0,1]$ and $\varepsilon > 0$, then there exists a polynomial function, p, such that* $\max|f(x) - p(x)| < \varepsilon$ *for* $0 \leqslant x \leqslant 1$.

Once you've seen this sequence of polynomials, the actual convergence proof is much less inspiring. However the use of some elementary probability theory can make the convergence proof a bit more interesting. This is not an original approach; it is actually a return to Bernstein's method of proof in [**1**]. Since this reference is not readily available, it is worth repeating.

The only tools that are needed are the binomial distribution and Chebyshev's Theorem. The binomial distribution is

$$b(n,x;j) = \binom{n}{j}x^j(1-x)^{n-j},$$

where n is the number of trials, x is the probability of success for any one trial, and j is the number of successes. Chebyshev's Theorem states that if a probability distribution has mean μ and standard deviation σ, the probability of obtaining a value that deviates from the mean by at least k standard deviations is at most $1/k^2$; i.e., $\Pr(|x-\mu| \geqslant k\sigma) \leqslant 1/k^2$.

To prove convergence of the Bernstein polynomials, define a sequence of families of random variables $F_{n,x}$, $0 \leqslant x \leqslant 1$ and $n = 0,1,2,\ldots$ in terms of a binomial random variable with n trials and probability of success x. If j successes occur, define the value of $F_{n,x}$ to be $f(j/n)$. Then the expected value of $F_{n,x}$ is

$$E(F_{n,x}) = \sum_{j=0}^{n} f\left(\frac{j}{n}\right)b(n,x;j) = B_n(f;x).$$

We now prove that $E(F_{n,x})$ converges uniformly to f on $[0,1]$ as $n \to \infty$.

Let $\varepsilon > 0$ be arbitrary. Since f is continuous, it is bounded on $[0,1]$; therefore there exists M such that $|f(x)| \leqslant M$ for $0 \leqslant x \leqslant 1$, which implies that $|f(x) - f(y)| \leqslant 2M$ for $0 \leqslant x \leqslant 1$. The continuity of f on $[0,1]$ also implies uniform continuity, i.e., there exists $\delta > 0$ such that if $|x-y| < \delta$, then $|f(x) - f(y)| < \varepsilon/2$. Now select a positive integer k such that $2M/k^2 < \varepsilon/2$ and a positive integer N such that $k/2\sqrt{N} < \delta$. Then if $n \geqslant N$ and $0 \leqslant x \leqslant 1$,

$$|f(x) - E(F_{n,x})| = \left|\sum_{j=0}^{n}\left(f(x) - f\left(\frac{j}{n}\right)\right)b(n,x;j)\right|$$

$$\leqslant \sum_{j=0}^{n} \left| f(x) - f\left(\frac{j}{n}\right) \right| b(n, x; j)$$

$$= \sum_{\left|\frac{j}{n} - x\right| < k/2\sqrt{n}} \left| f(x) - f\left(\frac{j}{n}\right) \right| b(n, x; j)$$

$$+ \sum_{\left|\frac{j}{n} - x\right| \geqslant k/2\sqrt{n}} \left| f(x) - f\left(\frac{j}{n}\right) \right| b(n, x; j)$$

$$\leqslant \sum_{\left|\frac{j}{n} - x\right| < \delta} \left| f(x) - f\left(\frac{j}{n}\right) \right| b(n, x; j) + 2M \Pr\left(\left| \frac{j}{n} - x \right| \geqslant k/2\sqrt{n} \right).$$

The first summand is less than $\varepsilon/2$, by our choice of δ. To bound the second summand, we use the fact that the mean and standard deviation of $b(n, x; k)$ are nx and $\sqrt{nx(1 - x)}$, respectively. So if

$$\left| \frac{j}{n} - x \right| \geqslant k/2\sqrt{n} \geqslant k\sqrt{\frac{x(1 - x)}{n}},$$

then $|j - nx| \geqslant k\sqrt{nx(1 - x)}$ and application of Chebyshev's Theorem yields

$$2M \sum_{\left|\frac{j}{n} - x\right| \geqslant k/2\sqrt{n}} b(n, x; j) \leqslant 2M/k^2 < \varepsilon/2.$$

This establishes that $|f(x) - B_n(f; x)| < \varepsilon$ for all $x \in [0,1]$ and $n \geqslant N$.

Reference

1. S. N. Bernstein, Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités, Comm. Soc. Math. Kharkov, 13 (1912) 1–2.

DUALITY IN THE CLASSROOM

GILBERT STRANG
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139

This note computes the smallest solution to an underdetermined system $Ax = b$. It is a problem that puts to use five or six different topics in linear algebra, and at the same time it illustrates one new idea: the magic of duality. Perhaps the simplest example of duality, in which we can see directly that a minimum equals a maximum, is the distance between a point and a line. The minimum distance from the origin to the points on a line (in R^3) equals the maximum distance to the planes through that line. That is evident to the intuition, which is extremely good at projection, and it is the special case of our problem when the matrix A is two by three.

In general A will be an m by n matrix with $m < n$; there are more unknowns than equations. We assume that the rows are independent, so the rank is m, AA^T is invertible, and $Ax = b$ has infinitely many solutions. The problem is to minimize the Euclidean length $|x| = (x_1^2 + \cdots + x_n^2)^{1/2}$, and it can be approached in at least four ways:

(i) by introducing Lagrange multipliers and then minimizing quadratics;
(ii) by developing the duality that was intuitive above;
(iii) as an instance of the pseudo-inverse A^+ (in this case a right inverse);
(iv) through projection.