Mathematical Tripos Part IB: Lent 2010

Numerical Analysis – Lecture 4[1]

2.4 Least-squares polynomial fitting

Given $f \in C[a,b]$ and a scalar product $\langle g,h\rangle = \int_a^b w(x)g(x)h(x)\,\mathrm{d}x$, we wish to pick $p \in \mathbb{P}_n[x]$ so as to minimise $\langle f-p, f-p\rangle$. Again, we stipulate that $w(x) > 0$ for $x \in (a,b)$. Intuitively speaking, p approximates f and is an alternative to an interpolating polynomial.
Let $p_0, p_1, \ldots, p_n$ be orthogonal polynomials w.r.t. the underlying inner product, $p_\ell \in \mathbb{P}_\ell[x]$. They form a basis of $\mathbb{P}_n[n]$, therefore for every $p \in \mathbb{P}_n$ there exist $c_0, c_1, \ldots, c_n \in \mathbb{R}$ such that $p = \sum_{k=0}^n c_k p_k$. Because of orthogonality,

$$\langle f-p, f-p\rangle = \left\langle f - \sum_{k=0}^n c_k p_k, f - \sum_{k=0}^n c_k p_k \right\rangle = \langle f,f\rangle - 2\sum_{k=0}^n c_k \langle p_k, f\rangle + \sum_{k=0}^n c_k^2 \langle p_k, p_k\rangle.$$

To derive optimal $c_0, c_1, \ldots, c_n$ we seek to minimise the last expression. (Note that it is a quadratic function in the c_is.) Since

$$\tfrac{1}{2}\frac{\partial}{\partial c_k}\langle f-p, f-p\rangle = -\langle p_k, f\rangle + c_k\langle p_k, p_k\rangle, \qquad k = 0,1,\ldots,n,$$

setting the gradient to zero yields

$$p(x) = \sum_{k=0}^n \frac{\langle p_k, f\rangle}{\langle p_k, p_k\rangle} p_k(x). \tag{2.3}$$

Note that

$$\langle f-p, f-p\rangle = \langle f,f\rangle - \sum_{k=0}^n \{2c_k\langle p_k, f\rangle - c_k^2\langle p_k, p_k\rangle\} = \langle f,f\rangle - \sum_{k=0}^n \frac{\langle p_k, f\rangle^2}{\langle p_k, p_k\rangle}. \tag{2.4}$$

This identity can be rewritten as $\langle f-p, f-p\rangle + \langle p,p\rangle = \langle f,f\rangle$, reminiscent of the Pythagoras theorem.

How to choose n? Note that $c_k = \langle p_k, f\rangle / \langle p_k, p_k\rangle$ is independent of n. Thus, we can continue to add terms to (2.3) until $\langle f-p, f-p\rangle$ is below specified *tolerance* ε. Because of (2.4), we need to pick n so that $\langle f,f\rangle - \varepsilon < \sum_{k=0}^n \langle p_k, f\rangle^2 / \langle p_k, p_k\rangle$.

Theorem *(The Parseval identity)* Let $[a,b]$ be finite. Then

$$\sum_{k=0}^\infty \frac{\langle p_k, f\rangle^2}{\langle p_k, p_k\rangle} = \langle f,f\rangle. \tag{2.5}$$

Incomplete proof. Let

$$\sigma_n := \sum_{k=0}^n \frac{\langle p_k, f\rangle^2}{\langle p_k, p_k\rangle}, \qquad n = 0,1,\ldots,$$

hence $\langle f-p, f-p\rangle = \langle f,f\rangle - \sigma_n \geq 0$. The sequence $\{\sigma\}_{n=0}^\infty$ increases monotonically and $\sigma_n \leq \langle f,f\rangle$ implies that $\lim_{n\to\infty}\sigma_n$ exists. According to the *Weierstrass theorem*, any function in $C[a,b]$ can be approximated arbitrarily close by a polynomial, hence $\lim_{n\to\infty}\langle f-p, f-p\rangle = 0$ and we deduce that $\sigma_n \stackrel{n\to\infty}{\longrightarrow} \langle f,f\rangle$ and (2.5) is true. $\square$

[1]Corrections and suggestions to these notes should be emailed to `A.Iserles@damtp.cam.ac.uk`. All handouts are available on the WWW at the URL `http://www.damtp.cam.ac.uk/user/na/PartIB/Handouts.html`.

2.5 Least-squares fitting to discrete function values

Suppose that $m \geq n+1$. We are given m function values $f(x_1), f(x_2), \ldots, f(x_m)$, where the x_ks are pairwise distinct, and seek $p \in \mathbb{P}_n[x]$ that minimises $\langle f-p, f-p \rangle$, where

$$\langle g, h \rangle := \sum_{k=1}^{m} g(x_k)h(x_k). \tag{2.6}$$

One alternative is to express p as $\sum_{\ell=0}^{n} p_\ell x^\ell$ and find optimal $p_0, \ldots, p_n$ using numerical linear algebra (which will be considered in the sequel). An alternative is to construct orthogonal polynomials w.r.t. the scalar product (2.6). The theory is identical to that of subsections 2.1–4, except that we have enough data to evaluate only $p_0, p_1, \ldots, p_{m-1}$. However, we need just $p_0, p_1, \ldots, p_n$ and $n \leq m-1$, and we have enough information to implement the algorithm. Thus

1. Employ the three-term recurrence (2.2) to calculate $p_0, p_1, \ldots, p_n$ (of course, using the scalar product (2.6));

2. Form $p(x) = \displaystyle\sum_{k=0}^{n} \frac{\langle p_k, f \rangle}{\langle p_k, p_k \rangle} p_k(x)$.

Since the work for each k is bounded by a constant multiple of m, the complete cost is $\mathcal{O}(mn)$, as compared with $\mathcal{O}\big(n^2 m\big)$ if linear algebra is used.

2.6 Gaussian quadrature

We are again in $C[a,b]$ and a scalar product is defined as in subsection 2.1, namely $\langle f, g \rangle = \int_a^b w(x)f(x)g(x)\,\mathrm{d}x$, where $w(x) > 0$ for $x \in (a,b)$. Our goal is to approximate integrals by finite sums,

$$\int_a^b w(x)f(x)\,\mathrm{d}x \approx \sum_{k=1}^{\nu} b_k f(c_k), \qquad f \in C[a,b].$$

The above is known as a *quadrature formula.* Here ν is given, whereas the points $b_1, \ldots, b_\nu$ (the *weights*) and $c_1, \ldots, c_\nu$ (the *nodes*) are independent of the choice of f.

A reasonable approach to achieving high accuracy is to require that the approximation is exact for all $f \in \mathbb{P}_m[x]$, where m is as large as possible – this results in *Gaussian quadrature* and we will demonstrate that $m = 2\nu - 1$ can be attained.

Firstly, we claim that $m = 2\nu$ is impossible. To prove this, choose arbitrary nodes $c_1, \ldots, c_\nu$ and note that $p(x) := \prod_{k=1}^{\nu}(x-c_k)^2$ lives in $\mathbb{P}_{2\nu}[x]$. But $\int_a^b w(x)p(x)\,\mathrm{d}x > 0$, while $\sum_{k=1}^{\nu} b_k p(c_k) = 0$ for any choice of weights $b_1, \ldots, b_\nu$. Hence the integral and the quadrature do not match.

Let $p_0, p_1, p_2, \ldots$ denote, as before, the monic polynomials which are orthogonal w.r.t. the underlying scalar product.

Theorem Given $n \geq 1$, all the zeros of p_n are real, distinct and lie in the interval (a,b).

Proof. Recall that $p_0 \equiv 1$. Thus, by orthogonality,

$$\int_a^b w(x)p_n(x)\,\mathrm{d}x = \int_a^b w(x)p_0(x)p_n(x)\,\mathrm{d}x = \langle p_0, p_n \rangle = 0$$

and we deduce that p_n changes sign at least once in (a,b).

Denote by $m \geq 1$ the number of the sign changes of p_n in (a,b) and assume that $m \leq n-1$. Denoting the points where a sign change occurs by $\xi_1, \xi_2, \ldots, \xi_m$, we let $q(x) := \prod_{j=1}^{m}(x-\xi_j)$. Since $q \in \mathbb{P}_m[x]$, $m \leq n-1$, it follows that $\langle q, p_n \rangle = 0$. On the other hand, it follows from our construction that $q(x)p_n(x)$ does not change sign throughout $[a,b]$ and vanishes at a finite number of points, hence

$$|\langle q, p_n \rangle| = \left| \int_a^b w(x)q(x)p_n(x)\,\mathrm{d}x \right| = \int_a^b w(x)|q(x)p_n(x)|\,\mathrm{d}x > 0,$$

a contradiction. It follows that $m = n$ and the proof is complete. $\square$