An interpolatory view of polynomial least squares approximation

Qingjun Chang* Chongyang Deng† Michael S. Floater‡

December 4, 2019

Abstract

We derive a formula for weighted polynomial least squares approximation which expresses the approximant as a convex combination of interpolants. There is a similar formula for L_2 approximation and the same principle applies to multivariate approximation.

Keywords: Polynomial least squares approximation, polynomial interpolation, Cauchy-Binet theorem.

1 Introduction

We usually think of least squares approximation as an alternative to interpolation. It is a way of reducing data and of avoiding the sometimes poor behaviour of interpolation due to the spacing of the points or noise in the data. In this paper we show that the two are nevertheless closely related. The approximant can be expressed as a weighted average of many interpolants, which are somehow smoothed out by this averaging.

Let $x_1, x_2, \ldots, x_m$, $m \geq 1$, be distinct points in $\mathbb{R}$, and let f be some real-valued function defined at these points. Let $w_1, w_2, \ldots, w_m > 0$ be positive weights associated with the points. As is well known, for any degree n,

*Hangzhou Dianzi University, Hangzhou, China, *email: qingjun.cn@gmail.com*

†Hangzhou Dianzi University, Hangzhou, China, *email: dcy@hdu.edu.cn*

‡Department of Mathematics, University of Oslo, Moltke Moes vei 35, 0851 Oslo, Norway, *email: michaelf@math.uio.no*

$0 \leq n \leq m-1$, the associated least squares approximation is the unique polynomial $p(x)$ of degree at most n that minimizes

$$\sum_{i=1}^{m} w_i(f(x_i) - p(x_i))^2. \tag{1}$$

One approach to finding p is to represent it in monomial form, and to find its coefficients by solving the normal equations or using QR factorization. Another is to construct orthogonal polynomials with respect to the points and weights, and then to find p as a linear combination of these. There are many texts on this topic, see for example, [2, Sec. 3.5] and [3, Chap. 5].

In this paper we derive a formula for p which expresses it as a weighted average of all the interpolants to f over sets of $n+1$ points among $x_1, \ldots, x_m$. Despite its simplicity, we have not been able to find it in the mathematical literature. The formula is 'combinatorial' and contains $\binom{m}{n+1}$ terms and so we are not suggesting to use it as a method of computing p. However, it is convenient for analyzing the influence of the weights on p. For example, as is well known, if we let the weights at one or more points tend to infinity, in the limit, p interpolates f at these points. One could find this limiting polynomial by imposing interpolation constraints at these points and minimizing over the remaining ones, using Lagrange multipliers. We instead obtain a formula as a weighted average of interpolants.

There is a similar formula for L_2 approximation and the same principle applies to multivariate approximation.

2 Interpolation formula

The interpolation formula is as follows. Let $\mathcal{P}_N$ denote the set of all subsets of $\{1, \ldots, m\}$ of cardinality N. The cardinality of $\mathcal{P}_N$ is $\binom{m}{N}$. For each $K \in \mathcal{P}_{n+1}$ let $p_K(x)$ denote the polynomial of degree at most n that interpolates f at the points x_i, $i \in K$, let w_K be the product of weights

$$w_K = \prod_{i \in K} w_i,$$

and let D_K be the product of squared distances

$$D_K = \prod_{\{i,j\} \subset K} (x_j - x_i)^2, \tag{2}$$

with the convention that this product is 1 if $n = 0$. Let $\lambda_K = w_K D_K$.

We will show

Theorem 1 *For $x \in \mathbb{R}$,*

$$p(x) = \sum_{K \in \mathcal{P}_{n+1}} \lambda_K p_K(x) \Big/ \sum_{K \in \mathcal{P}_{n+1}} \lambda_K. \tag{3}$$

For example, in the linear case, the theorem tells us that

$$p(x) = \sum_{\{i,j\} \in \mathcal{P}_2} \lambda_{ij} p_{ij}(x) \Big/ \sum_{\{i,j\} \in \mathcal{P}_2} \lambda_{ij},$$

where p_{ij} is the linear interpolant to f at x_i and x_j and

$$\lambda_{ij} = w_i w_j (x_j - x_i)^2. \tag{4}$$

In the quadratic case,

$$p(x) = \sum_{\{i,j,k\} \in \mathcal{P}_3} \lambda_{ijk} p_{ijk}(x) \Big/ \sum_{\{i,j,k\} \in \mathcal{P}_3} \lambda_{ijk},$$

where

$$\lambda_{ijk} = w_i w_j w_k (x_j - x_i)^2 (x_k - x_i)^2 (x_k - x_j)^2, \tag{5}$$

and p_{ijk} is the quadratic interpolant to f at x_i, x_j, x_k. At the other extreme, if $n = m - 1$, p interpolates f. If $n = m - 2$, then

$$p(x) = \sum_{i=1}^{m} \widehat{\lambda}_i p_{\{1,\ldots,m\} \setminus i} \Big/ \sum_{i=1}^{m} \widehat{\lambda}_i,$$

where, by cancelling common factors in $\lambda_{\{1,\ldots,m\} \setminus i}$,

$$\widehat{\lambda}_i = \frac{1}{w_i} \prod_{j \neq i} \frac{1}{(x_j - x_i)^2}.$$

Proof. Let

$$p(x) = \sum_{j=0}^{n} c_j x^j. \tag{6}$$

be the polynomial that minimizes (1). The normal equations for (1) can be written as $A\mathbf{c} = \mathbf{b}$, where

$$A = \begin{bmatrix} \sum w_k & \sum w_k x_k & \cdots & \sum w_k x_k^n \\ \sum w_k x_k & \sum w_k x_k^2 & \cdots & \sum w_k x_k^{n+1} \\ \vdots & \vdots & & \vdots \\ \sum w_k x_k^n & \sum w_k x_k^{n+1} & \cdots & \sum w_k x_k^{2n} \end{bmatrix},$$

$\mathbf{c} = (c_0, c_1, \ldots, c_n)^T$, and

$$\mathbf{b} = \left(\sum w_k f(x_k), \sum w_k x_k f(x_k), \ldots, \sum w_k x_k^n f(x_k) \right)^T,$$

and all summations are over $k = 1, \ldots, m$. By Cramer's rule, c_j is the quotient of determinants $c_j = |A_j|/|A|$, $j = 0, 1, \ldots, n$, where A is as above and A_j is the matrix formed by replacing the $(j+1)$-st column of A by $\mathbf{b}$.

We now expand the determinants of A and A_j, $j = 0, 1, \ldots, n$. Firstly, we can write A as the product

$$A = \begin{bmatrix} w_1 & \cdots & w_m \\ w_1 x_1 & \cdots & w_m x_m \\ \vdots & & \vdots \\ w_1 x_1^n & \cdots & w_m x_m^n \end{bmatrix} \begin{bmatrix} 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \cdots & x_m^n \end{bmatrix}.$$

Then, by the Cauchy-Binet theorem [4, p. 1],

$$|A| = \sum_{K \in \mathcal{P}_{n+1}} w_K |V_K|^2, \tag{7}$$

where V_K is the Vandermonde matrix

$$V_K = \begin{bmatrix} 1 & x_{k_0} & \cdots & x_{k_0}^n \\ 1 & x_{k_1} & \cdots & x_{k_1}^n \\ \vdots & & & \vdots \\ 1 & x_{k_n} & \cdots & x_{k_n}^n \end{bmatrix},$$

and $K = \{k_0, k_1, \ldots, k_n\}$ with $k_0 < k_1 < \cdots < k_n$. As is well known, the determinant of V_K is

$$|V_K| = \prod_{0 \le \alpha < \beta \le n} (x_{k_\beta} - x_{k_\alpha}),$$

and so its square is D_K in (2). While the sign of $|V_K|$ depends on the ordering of $k_0, k_1, \ldots, k_n$, D_K does not. Secondly, we can similarly express A_j as the product

$$A_j = \begin{bmatrix} w_1 & \cdots & w_m \\ w_1 x_1 & \cdots & w_m x_m \\ \vdots & & \vdots \\ w_1 x_1^n & \cdots & w_m x_m^n \end{bmatrix} \begin{bmatrix} 1 & \cdots & x_1^{j-1} & f(x_1) & x_1^{j+1} & \cdots & x_1^n \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & \cdots & x_m^{j-1} & f(x_m) & x_m^{j+1} & \cdots & x_m^n \end{bmatrix},$$

and using the Cauchy-Binet theorem again,

$$|A_j| = \sum_{K \in \mathcal{P}_{n+1}} w_K |V_K||V_{j,K}|, \tag{8}$$

where

$$V_{j,K} = \begin{bmatrix} 1 & \cdots & x_{k_0}^{j-1} & f(x_{k_0}) & x_{k_0}^{j+1} & \cdots & x_{k_0}^n \\ \vdots & & \vdots & \vdots & & \vdots & \\ 1 & \cdots & x_{k_n}^{j-1} & f(x_{k_n}) & x_{k_n}^{j+1} & \cdots & x_{k_n}^n \end{bmatrix}.$$

Next observe that the polynomial interpolant p_K to f can also be written in monomial form, as

$$p_K(x) = \sum_{j=0}^{n} c_{j,K} x^j,$$

and the interpolation conditions give the linear system $V_K \mathbf{c}_K = \mathbf{f}_K$, where $\mathbf{c}_K = (c_{0,K}, c_{1,K}, \ldots, c_{n,K})^T$, and $\mathbf{f}_K = (f(x_{k_0}), f(x_{k_1}), \ldots, f(x_{k_n}))^T$. By Cramer's rule applied to this linear system, we have $c_{j,K} = |V_{j,K}|/|V_K|$. Therefore,

$$\begin{aligned} p(x) = \sum_{j=0}^{n} c_j x^j &= \sum_{j=0}^{n} \frac{\sum_K w_K |V_K||V_{j,K}| x^j}{\sum_K w_K |V_K|^2} \\ &= \sum_{j=0}^{n} \frac{\sum_K w_K |V_K|^2 c_{j,K} x^j}{\sum_K w_K |V_K|^2} = \frac{\sum_K w_K |V_K|^2 p_K(x)}{\sum_K w_K |V_K|^2}, \end{aligned}$$

as claimed. □

Figure 1 shows an implementation of the formula of Theorem 1, using degree $n = 1$ on the left and degree $n = 5$ on the right. The weights here are equal to 1.

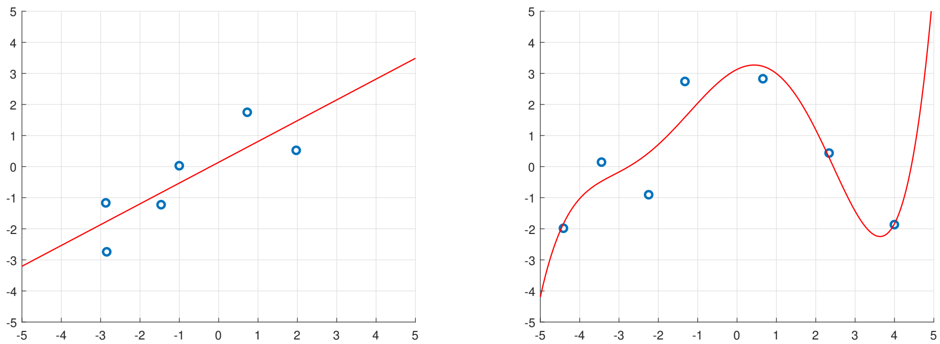

Figure 1: Least squares using degrees $n = 1$ and $n = 5$.

3 Some simple consequences

For any $r = 0, 1, \ldots, n$, the r-th derivative of p is also the same weighted average of the r-th derivatives of the interpolants,

$$p^{(r)}(x) = \sum_{K \in \mathcal{P}_{n+1}} \lambda_K p_K^{(r)}(x) \Big/ \sum_{K \in \mathcal{P}_{n+1}} \lambda_K, \tag{9}$$

and so we have the simple upper and lower bounds

$$\min_{K \in \mathcal{P}_{n+1}} p_K^{(r)}(x) \le p^{(r)}(x) \le \max_{K \in \mathcal{P}_{n+1}} p_K^{(r)}(x).$$

Figure 2 shows an example of the upper and lower bounds when $n = 1$ and $r = 0$.

Similarly, the approximation error shares the same weighted average,

$$f(x) - p(x) = \sum_{K \in \mathcal{P}_{n+1}} \lambda_K (f(x) - p_K(x)) \Big/ \sum_{K \in \mathcal{P}_{n+1}} \lambda_K,$$

and so

$$f(x) - p(x) = \sum_{K \in \mathcal{P}_{n+1}} \lambda_K \phi_K(x) [\{x_i : i \in K\}, x] f \Big/ \sum_{K \in \mathcal{P}_{n+1}} \lambda_K,$$

where $[\{x_i : i \in K\}, x]f$ is the divided difference of f at the points x_i, $i \in K$, and x, and

$$\phi_K(x) := \prod_{i \in K} (x - x_i).$$

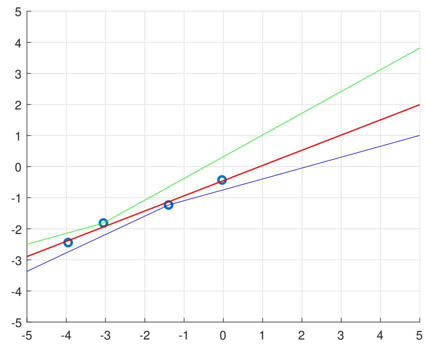

Figure 2: Upper and lower bounds with $n = 1$.

4 Influence of the weights

It is clear from the interpolation formula that if we increase the weight w_i for some $i \in \{1, \dots, m\}$, the interpolants p_K for which $i \in K$ will have a greater influence on p than the others. In the limiting case that $w_i \to \infty$, while keeping the other weights fixed, we expect that p will interpolate f at x_i. Similarly, for any distinct i, j, if $w_i, w_j \to \infty$, we expect that p will interpolate f at both x_i and x_j. We now confirm this mathematically and derive a formula.

Let I be some subset of $\{1, \dots, m\}$ of cardinality r, where $1 \le r \le n+1$, and let $q_I(x)$ denote the limit of $p(x)$ as $w_i \to \infty$ for all $i \in I$. Let $\mathcal{P}_N(I)$ denote the set of all subsets of $\{1, \dots, m\} \setminus I$ of cardinality N. The cardinality of $\mathcal{P}_N(I)$ is $\binom{m-r}{N}$.

Corollary 1 *If $r = n + 1$ then $q_I = p_I$ while if $r < n + 1$,*

$$q_I(x) = \sum_{J \in \mathcal{P}_{n+1-r}(I)} \mu_{I,J} p_{I \cup J}(x) \Big/ \sum_{J \in \mathcal{P}_{n+1-r}(I)} \mu_{I,J}. \tag{10}$$

where $\mu_{I,J} = w_J D_{I \cup J}$.

Clearly, in both cases $q_I(x_i) = f(x_i)$ for all $i \in I$.

Proof. We split the sums in (3) in both the numerator and the denominator into two:

$$\sum_K = \sum_{K: I \subset K} + \sum_{K: I \not\subset K}.$$

We then divide both numerator and denominator by w_I and use the fact that when $I \subset K$, $w_K/w_I = w_{K\setminus I}$. Then letting $w_i \to \infty$ for $i \in I$, $p(x)$ converges to the limit

$$q_I(x) = \frac{\sum_{K:I\subset K} w_{K\setminus I} D_K p_K(x)}{\sum_{K:I\subset K} w_{K\setminus I} D_K}.$$

If $r = n+1$ we must have $K = I$ in the two sums and we obtain $q_I(x) = p_I(x)$. If $r < n+1$ we express K as the disjoint union $I \cup J$ and change the sum over K to a sum over J. $\square$

For example, in the linear case, the limit of $p(x)$ as $w_i \to \infty$ is

$$q_i(x) = \sum_{j\neq i} \mu_{ij} p_{ij}(x) \Big/ \sum_{j\neq i} \mu_{ij},$$

where

$$\mu_{ij} = w_j(x_j - x_i)^2. \tag{11}$$

In the quadratic case, the limit of $p(x)$ as $w_i \to \infty$ is

$$q_i(x) = \sum_{\{j,k\}\in\mathcal{P}_2(i)} \mu_{ijk} p_{ijk}(x) \Big/ \sum_{\{j,k\}\in\mathcal{P}_2(i)} \mu_{ijk},$$

where

$$\mu_{ijk} = w_j w_k (x_j - x_i)^2 (x_k - x_i)^2 (x_k - x_j)^2. \tag{12}$$

If the cardinality of I is greater than 1, there is some cancellation in the limit formula (10) because

$$D_{I\cup J} = D_I D_{I,J} D_J$$

where

$$D_{I,J} = \prod_{i\in I, j\in J} (x_j - x_i)^2,$$

and then we can replace $\mu_{I,J}$ in (10) by $w_J D_{I,J} D_J$. As an example of this, in the quadratic case, the limit of $p(x)$ as $w_i, w_j \to \infty$, $i \neq j$, is

$$q_{ij}(x) = \sum_{k\neq i,j} \mu_{ijk} p_{ijk}(x) \Big/ \sum_{k\neq i,j} \mu_{ijk} p_{ijk}(x),$$

where

$$\mu_{ijk} = w_k (x_k - x_i)^2 (x_k - x_j)^2.$$

Figure 3 shows the result of applying the formula of Corollary 1 using the same data as in Figure 1 but with $w_3 = \infty$ on the left and $w_2, w_4, w_6 = \infty$ on the right. All other weights are equal to 1.

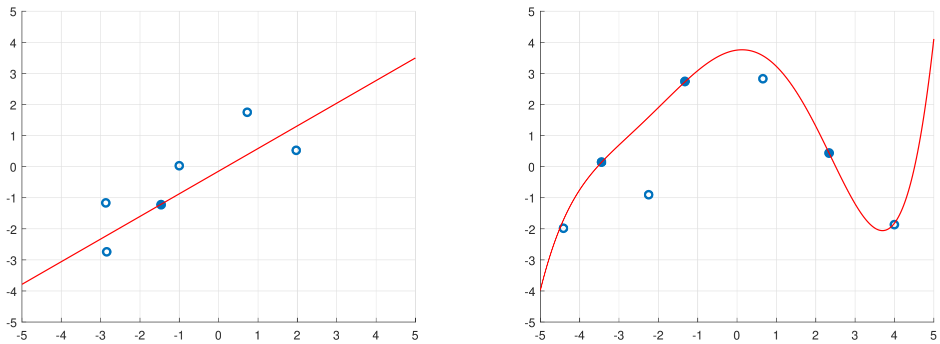

Figure 3: Limiting case when some weights are set to infinity.

5 Derivative estimation

Least squares approximation is often used to estimate derivatives. From (9),

$$f^{(r)}(x) \approx p^{(r)}(x) = \sum_{K \in \mathcal{P}_{n+1}} \lambda_K p_K^{(r)}(x) \Big/ \sum_{K \in \mathcal{P}_{n+1}} \lambda_K,$$

for $r = 1, \ldots, n$. If we want to estimate $f^{(r)}$ at some point x_i and we trust the value of f there we might prefer to let $w_i \to \infty$ and take the r-th derivative of the limiting polynomial. Then Corollary 1 gives

$$f^{(r)}(x_i) \approx q_i^{(r)}(x_i) = \sum_J \mu_{i,J} p_{i \cup J}^{(r)}(x_i) \Big/ \sum_J \mu_{i,J}.$$

In the linear case this gives the first derivative estimate

$$f'(x_i) \approx \sum_{j \neq i} \mu_{ij} [x_i, x_j] f \Big/ \sum_{j \neq i} \mu_{ij}.$$

with μ_{ij} as in (11).

6 L_2 approximation

There is a similar interpolation formula for L_2 approximation in an interval $[a, b]$. Suppose $f : [a, b] \to \mathbb{R}$ is a continuous function and $w : [a, b] \to \mathbb{R}$

is a positive, integrable weight function. For any $n \geq 0$, there is a unique polynomial p of degree at most n that minimizes

$$\int_a^b w(t)(f(t) - p(t))^2 \, dt. \tag{13}$$

To describe the formula let $\mathcal{Q}_{n+1}$ denote the set of all sets of $n+1$ distinct points in $[a, b]$. Each $T \in \mathcal{Q}_{n+1}$ has the form

$$T = (t_0, t_1, \ldots, t_n), \qquad a \leq t_0 < t_1 < \cdots < t_n \leq b.$$

Let $p_T(x)$ be the polynomial interpolant to f of degree at most n at the points of T, let

$$w_T = \prod_{i=0}^{n} w(t_i),$$

let

$$D_T = \prod_{0 \leq i < j \leq n} (t_j - t_i)^2,$$

and set $\lambda_T = w_T D_T$. Further, define

$$\int_{T \in \mathcal{Q}_{n+1}} F = \int_{a \leq t_0 < t_1 < \cdots < t_n \leq b} F(t_0, t_1, \ldots, t_n) \, dt_0 dt_1 \cdots dt_n,$$

for a function F of $n+1$ variables, integrable in this sense.

We claim

Theorem 2 *For $x \in \mathbb{R}$,*

$$p(x) = \int_{T \in \mathcal{Q}_{n+1}} \lambda_T p_T(x) \Big/ \int_{T \in \mathcal{Q}_{n+1}} \lambda_T. \tag{14}$$

For example, in the linear case, the theorem says that

$$p(x) = \int_{a \leq s < t \leq b} \lambda_{st} p_{st}(x) \, ds \, dt \Big/ \int_{a \leq s < t \leq b} \lambda_{st} \, ds \, dt,$$

where p_{st} is the linear interpolant to f at the points s and t, and

$$\lambda_{st} = w(s) w(t) (t - s)^2.$$

We note that $p_T(x)$ regarded as a function of T with x fixed is continuous in T if $f \in C^n[a,b]$. In the limit as two points in T approach each other, p_T becomes a Hermite interpolant; see [1] and [3]. However, even for f merely continuous, the integral in the numerator of (14) is well defined due to the term D_T. Writing p_T in Lagrange form

$$p_T(x) = \sum_{i=0}^{n} L_{i,T}(x) f(s), \qquad L_{i,T}(x) = \prod_{j=0, j\neq i}^{n} \frac{x - t_j}{t_i - t_j},$$

we see that the division by the differences $t_i - t_j$ is cancelled out by the differences in D_T. For each $i = 0, \dots, n$, the product $D_T L_{i,T}(x)$ is a polynomial in $t_0, \dots, t_n$, and therefore the product $w_T D_T L_{i,T}(x) f(t_i)$ is integrable with respect to T.

Proof. The proof is analogous to the discrete case. We write p as in (6). The normal equations for the minimization of (13) are $A\mathbf{c} = \mathbf{b}$, where

$$A = \begin{bmatrix} \int w(t) & \int w(t)t & \cdots & \int w(t)t^n \\ \int w(t)t & \int w(t)t^2 & \cdots & \int w(t)t^{n+1} \\ \vdots & \vdots & & \vdots \\ \int w(t)t^n & \int w(t)t^{n+1} & \cdots & \int w(t)t^{2n} \end{bmatrix},$$

$\mathbf{c} = (c_0, c_1, \dots, c_n)^T$, and

$$\mathbf{b} = \left(\int w(t)f(t), \int w(t)tf(t), \dots, \int w(t)t^n f(t) \right)^T,$$

and all integrals are over t in $[a,b]$. Using Cramer's rule, $c_j = |A_j|/|A|$, $j = 0, 1, \dots, n$, with A as above and A_j the matrix formed by replacing the $(j+1)$-st column of A by $\mathbf{b}$.

To find the determinant of A, we can write it as

$$A = \left[\int_a^b \phi_i(t)\psi_k(t)\, dt \right]_{i,k=0,1,\dots,n},$$

where $\phi_i(t) = w(t)t^i$, and $\psi_i(t) = t^i$. Then by the integral version of the Cauchy-Binet theorem [4, Formula (2.5)],

$$|A| = \int_{T \in \mathcal{Q}_{n+1}} |[\phi_k(t_i)]_{ik}||[\psi_k(t_i)]_{ik}|,$$

and therefore,

$$|A| = \int_{T \in \mathcal{Q}_{n+1}} w_T |V_T|^2,$$

where V_T is the Vandermonde determinant

$$V_T = \begin{bmatrix} 1 & t_0 & \cdots & t_0^n \\ 1 & t_1 & \cdots & t_1^n \\ \vdots & & & \vdots \\ 1 & t_n & \cdots & t_n^n \end{bmatrix}.$$

To find $|A_j|$ we write A_j as

$$A_j = \left[\int_a^b \phi_i(t) \hat{\psi}_k(t)\, dt \right]_{i,k=0,1,\ldots,n},$$

where $\hat{\psi}_i(t) = t^i$, $i \neq j$, and $\hat{\psi}_j(t) = f(t)$.

Then the integral version of the Cauchy-Binet theorem gives

$$|A_j| = \int_{T \in \mathcal{Q}_{n+1}} |[\phi_k(t_i)]_{ik}||[\hat{\psi}_k(t_i)]_{ik}|,$$

and therefore,

$$|A_j| = \int_{T \in \mathcal{Q}_{n+1}} w_T |V_T||V_{j,T}|,$$

where

$$V_{j,T} = \begin{bmatrix} 1 & \cdots & t_0^{j-1} & f(t_0) & t_0^{j+1} & \cdots & t_0^n \\ \vdots & & \vdots & \vdots & & \vdots & \\ 1 & \cdots & t_n^{j-1} & f(t_n) & t_n^{j+1} & \cdots & t_n^n \end{bmatrix}.$$

Similar to the discrete case,

$$p_T(x) = \sum_{j=0}^{n} c_{j,T} x^j,$$

where $c_{j,T} = |V_{j,T}|/|V_T|$, and the remaining steps of the proof are the same as in the proof of Theorem 1. $\square$

7 Multivariate approximation

The same principle applies in the multivariate case. Let us just illustrate this in the bivariate case. Suppose $(x_1, y_1), \ldots, (x_m, y_m)$ are points in $\mathbb{R}^2$. Given weights $w_i > 0$ and function values $f(x_i, y_i)$, $i = 1, \ldots, m$, consider the problem of finding a polynomial $p(x, y)$ of degree at most n that minimizes

$$\sum_{i=1}^{m} w_i (f(x_i, y_i) - p(x_i, y_i))^2.$$

Letting $N = \binom{n+2}{2}$, we can represent p as

$$p(x, y) = \sum_{j=1}^{N} c_j B_j(x, y),$$

with respect to the basis

$$(B_1(x, y), \ldots, B_N(x, y)) = (1, x, y, x^2, xy, y^2, \ldots, x^n, x^{n-1}y, \ldots, y^n),$$

and we will assume that $N \le m$. Let $V \in \mathbb{R}^{m,N}$ be the matrix

$$V = [B_j(x_i, y_i)]_{i=1,\ldots,m, j=1,\ldots,N},$$

and for each $K = \{k_1, k_2, \ldots, k_N\} \in \mathcal{P}_N$ with $k_1 < k_2 < \cdots < k_N$, let V_K be the square submatrix

$$V_K = [B_j(x_{k_i}, y_{k_i})]_{i,j=1,\ldots,N}.$$

Let us next suppose that the points (x_i, y_i) are such that there is at least one $K \in \mathcal{P}_N$ such that V_K is non-singular. Then V has full rank N, and, as is well known, there is a unique minimizer p.

But we can now go a step further and derive a formula for p in terms of interpolants. Letting

$$\widehat{\mathcal{P}}_N = \{K \in \mathcal{P}_N : |V_K| \neq 0\},$$

and following the steps of the proof of Theorem 1 we obtain the formula

$$p(x, y) = \sum_{K \in \widehat{\mathcal{P}}_N} \lambda_K p_K(x, y) \Big/ \sum_{K \in \widehat{\mathcal{P}}_N} \lambda_K, \qquad (x, y) \in \mathbb{R}^2. \tag{15}$$

Here, similar to before, p_K is the polynomial of degree at most n that interpolates f at the points (x_i, y_i), $i \in K$, $w_K = \prod_{i \in K} w_i$, and $\lambda_K = w_K |V_K|^2$. However, the sum is only over subsets K for which $|V_K| \neq 0$ (and p_K is well defined). This is because when we follow the proof of Theorem 1 to derive (15), and we reach the two sums analogous to (7) and (8), any K for which $|V_K| = 0$ can be discarded.

For example, in the linear case $n = 1$, we have $N = 3$ and for $K = \{k_1, k_2, k_3\}$ in $\mathcal{P}_3$ with $k_1 < k_2 < k_3$, we have

$$V_K = \begin{bmatrix} 1 & x_{k_1} & y_{k_1} \\ 1 & x_{k_2} & y_{k_2} \\ 1 & x_{k_3} & y_{k_3} \end{bmatrix}.$$

Thus the sum in (15) is over triples of points that are not collinear.

Similar to the univariate case, partial derivatives of p can be expressed as averages of the partial derivatives of the p_K. The same principle applies to approximation in several variables.

References

[1] S. D. Conte and C. de Boor, *Elementary Numerical Analysis*, McGraw-Hill, 1980.

[2] A. C. Faul, *A Concise Introduction to Numerical Analysis*, Taylor and Francis, Florida, 2016.

[3] E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*, Dover, 1994.

[4] S. Karlin, *Total Positivity, Vol. I*, Stanford University Press, Stanford, California, 1968.