Least-Squares Fitting Using Orthogonal Multinomials

RICHARD H. BARTELS
University of Waterloo
and
JOHN J. JEZIORANSKI
Ontario Cancer Institute

Forsythe [3] has given a method for generating basis polynomials in a single variable that are orthogonal with respect to a given inner product. Weisfeld [6] later demonstrated that Forsythe's approach could be extended to polynomials in an arbitrary number of variables. In this paper we sharpen Weisfeld's results and present a method for computing weighted, multinomial, least-squares approximations to given data.

Categories and Subject Descripters: G.1.9 [**Numerical Analysis**]: Optimization—*least squares methods*; F.2.1 [**Analysis of Algorithms and Problem Complexity**]: Numerical Algorithms and Problems—*computation of polynomials*

General Terms: Algorithms

Additional Key Words and Phrases: Data fitting, multinomials, orthogonal, regression

The Algorithm: CONSTR and EVAL: Routines for Fitting Multinomials in a Least-Squares Sense. *ACM Trans. Math. Softw. 11*, 3 (September, 1985), 218–228.

1. INTRODUCTION, REVIEW, AND PRELIMINARY NOTATION

Let $\langle , \rangle$ denote an inner product on an appropriate class of real-valued functions of n real variables. Let $\psi_1, \ldots, \psi_m$ be linearly independent functions, and let $F(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) + e(x_1, \ldots, x_n)$, where e represents some random error. The problem of fitting f, in the least-squares sense, in the space spanned by the ψ_i requires that coefficients $c = c_1, \ldots, c_m$ be determined so that

$$P(x_1, \ldots, x_n; c_1, \ldots, c_m) = \sum_{i=1}^{m} c_i \psi_i(x_1, \ldots, x_n)$$

minimizes

$$\langle F - P, F - P \rangle.$$

This research was supported under Grant A4076 by the National Sciences and Engineering Research Council of Canada and by the Laidlaw Corporation under a fellowship administered by the University of Waterloo Computer Science Department.
Authors' addresses: Richard H. Bartels, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G2; John J. Jezioranski, Ontario Cancer Institute, 500 Sherbourne St., Toronto, Ontario, Canada M4X 1K9.

The coefficients $c_1, \ldots, c_m$ will be given by the solution to the normal equations

$$\begin{bmatrix} \langle \psi_1, \psi_1 \rangle & \cdots & \langle \psi_1, \psi_m \rangle \\ \vdots & & \vdots \\ \langle \psi_m, \psi_1 \rangle & \cdots & \langle \psi_m, \psi_m \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \langle \psi_1, F \rangle \\ \vdots \\ \langle \psi_m, F \rangle \end{bmatrix}, \tag{1.1}$$

provided that these equations are nonsingular.

For multinomial least-squares fitting $\psi_1, \ldots, \psi_m$ are often taken to be the first m monomials in the n variables of the problem, ψ's being arranged according to some increasing power order. For example,

$$n = 1,$$

$$\langle g, h \rangle = \sum_{i=1}^{N} w^{(i)} g(x^{(i)}) h(x^{(i)});$$

$$\psi_1(x), \ldots, \psi_m(x) \equiv 1, \quad x, x^2, \ldots, x^{m-1}$$

or

$$n = 2,$$

$$\langle g, h \rangle = \int_0^1 \int_0^1 w(x_1, x_2) g(x_1, x_2) h(x_1, x_2)\, dx_1\, dx_2,$$

$$\psi_1(x_1, x_2), \ldots, \psi_m(x_1, x_2) \equiv 1, \quad x_1, x_2, x_1^2, x_1 x_2, x_2^2, \ldots, x_1^{m_1} x_2^{m_2}.$$

A choice of this sort of basis for multinomials will generally result in an ill-conditioned system of equations (1.1). If, instead of the monomials, we choose a basis of multinomials that are orthogonal with respect to $\langle\ ,\ \rangle$, then the matrix of (1.1) will be diagonal.

One may, of course, use the Gram–Schmidt process to orthogonalize the basis $\psi_1, \ldots, \psi_m$, but Forsythe [3] and Weisfeld [6] have shown that a revised Gram–Schmidt process is more efficient. Forsythe worked with

$$n = 1,$$

$$\langle g, h \rangle = \sum_{i=1}^{N} w^{(i)} g(x^{(i)}) h(x^{(i)})$$

and generated $\psi_1, \ldots, \psi_m$ by

$$\psi_1 = 1,$$

$$\psi_j = x\psi_{j-1} - \sum_{i<j} \alpha_{j,i} \psi_i.$$

This construction ensures that ψ_j will be monic with highest-order term equal to x^{j-1}. (A normalizing constant is often added to the above, which we shall ignore for the sake of simplicity.)

It is easily verified that ψ_j will be orthogonal to $\psi_{j-1}, \ldots, \psi_1$ *if* $\alpha_{j,i}$ is chosen as

$$\alpha_{j,i} = \frac{\langle x\psi_{j-1}, \psi_i \rangle}{\langle \psi_i, \psi_i \rangle}$$

for all $l = 1, \ldots, j-1$. Forsythe showed that $\langle x\psi_{j-1}, \psi_l \rangle = 0$ for all $j - l > 2$ (i.e., $l < j - 2$). That is, the construction of the orthogonal basis of multinomials is accomplished by a three-term recurrence.

Weisfeld considered general inner products and a general n. He introduced the notation of vector indices

$$\psi_J = \psi_{(j_1, \ldots, j_n)},$$

where

$$J = (j_1, \ldots, j_n).$$

By letting $\sigma(J) = j_1 + \cdots + j_n$ for any vector index J, he defined

$$I < J$$

to mean

$$\sigma(I) < \sigma(J), \tag{1.2.1}$$

or else

$$\sigma(I) = \sigma(J) \tag{1.2.2a}$$

and

$$i_l + \cdots + i_n < j_l + \cdots + j_n \tag{1.2.2b}$$

for some $l \leq n$. (Note that this does not define a unique ordering. Two distinct sequences, among many possible ones, are produced by consistently choosing l in the above to be the *least*, respectively *greatest*, integer such that $l \leq n$.)

Given a vector index

$$J = (j_1, \ldots, j_k, 0, \ldots, 0), \tag{1.3}$$

where $j_k > 0$, Weisfeld defined an associated vector index

$$\hat{J} = (j_1, \ldots, j_k - 1, 0, \ldots, 0), \tag{1.4}$$

and constructed

$$\begin{aligned} \psi_{(0,\ldots,0)} &\equiv 1, \\ \psi_J &= x_k \psi_{\hat{J}} - \sum_{L<J} \alpha_{J,L} \psi_L. \end{aligned} \tag{1.5}$$

As before, it is easily verified that ψ_J will be orthogonal to all ψ_L $(L < J)$ if $\alpha_{J,L}$ is chosen as

$$\alpha_{J,L} = \frac{\langle x_k \psi_{\hat{J}}, \psi_L \rangle}{\langle \psi_L, \psi_L \rangle}.$$

Weisfeld proved that $\alpha_{J,L}$ is zero for all $L < J$ such that $\sigma(J) - \sigma(L) > 2$. This constitutes a generalization to n variables of the three-term recurrence.

In this paper we improve upon Weisfeld's results by taking greater care in fixing the ordering. We shall also concern ourselves with the details of mapping vector indices onto the natural numbers to facilitate transcribing our results into a computer program. We will find that more coefficients α turn out to be zero than are indicated in Weisfeld's results, and we are able to take advantage of some identities among the α's.

2. ORDERING

We need to establish a specific mapping from the monomials to the integers, one that will provide a convenient framework for dealing with the orthogonal multinomials. To do this, we arrange the monomials involving n variables in a triangular pattern in which the rth row contains all monomials of $(r-1)$st power, and each row of which past the first is organized into n ranges:

Definition 1.
Row 1 contains only 1.
Row 2 contains the n ranges $x_1, x_2, \ldots, x_n$; that is, the kth range contains only the monomial x_k.
Row r has as its kth range the monomials found by multiplying x_k by each member, in order, of ranges $k, \ldots, n$ in row $r-1$.
For example, when $n = 3$,

Row 1: 1
Row 2: $x_1 \quad x_2 \quad x_3$
Row 3: $x_1^2 \quad x_1x_2 \quad x_1x_3 \quad x_2^2 \quad x_2x_3 \quad x_3^2$
Row 4: $x_1^3 \quad x_1^2x_2 \quad x_1^2x_3 \quad x_1x_2^2 \quad x_1x_2x_3 \quad x_1x_3^2 \quad x_2^3 \quad x_2^2x_3 \quad x_2x_3^2 \quad x_3^3$

The three ranges in Row 3 are

$$x_1 \times \{x_1 \quad x_2 \quad x_3\} \equiv \{x_1^2 \quad x_1x_2 \quad x_1x_3\},$$

$$x_2 \times \{x_2 \quad x_3\} \equiv \{x_2^2 \quad x_2x_3\},$$

$$x_3 \times \{x_3\} \equiv \{x_3^2\}.$$

Our notational conventions for this table will be as follows:

(1) The position in the table of current interest will be indexed by j.
(2) The jth monomial in the table will be denoted by $\mu(j)$.
(3) The vector of exponents associated with $\mu(j)$ will be denoted by $\nu(j)$.
(4) $\sigma(j)$ is the sum of the exponents associated with $\mu(j)$, that is, the sum of the components of $\nu(j)$.

The monomials $\mu(j)$, their position j in the table, and their exponent vectors $\nu(j)$ are all ordered sets, all isomorphic to each other. The $n = 3$ example gives

$$\{j\} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, \ldots\},$$

$$\{\mu(j)\} = \{1, x_1, x_2, x_3, x_1^2, x_1x_2, x_1x_3, x_2^2, x_2x_3, \ldots\},$$

$$\{\nu(j)\} = \{(0,0,0), (1,0,0), (0,1,0), (0,0,1), (2,0,0), (1,1,0),$$
$$(1,0,1), (0,2,0), (0,1,1), \ldots\}.$$

We shall use the symbol $<$ for each of these three sets with the meanings

(1) $j^* < j^{**} \equiv j^*$ comes before j^{**} in $\{j\}$;
(2) $\nu(j^*) < \nu(j^{**}) \equiv (j_1^*, \ldots, j_n^*)$ comes before $(j_1^{**}, \ldots, j_n^{**})$ in $\{\nu(j)\}$;
(3) $\mu(j^*) < \mu(j^{**}) \equiv x_1^{j_1*} \cdots x_n^{j_n*}$ comes before $x_1^{j_1**} \cdots x_n^{j_n**}$ in $\{\mu(j)\}$.

The symbol $\leq$ will denote

$$j^* < j^{**} \qquad \text{or} \qquad j^* = j^{**},$$

and similarly for $\nu(j)$ and $\mu(j)$.

With respect to our notation, a monomial $\mu(j)$ associated with

$$\nu(j) = (0, \ldots, 0, j_{k_j}, \ldots, j_n), \qquad j_{k_j} > 0,$$

is in the k_jth range of the row numbered $(j_{k_j} + \cdots + j_n + 1) = \sigma(j) + 1$, and this monomial was constructed by multiplying the monomial associated with the exponent vector

$$(0, \ldots, 0, [j_{k_j} - 1], \ldots, j_n)$$

by x_{k_j}.

By convention, if

$$\nu(j) = (0, \ldots, 0, j_{k_j}, \ldots, j_n), \qquad j_{k_j} > 0,$$

we will define j' to be the index in the table for which

$$\nu(j') = (0, \ldots, 0, [j_{k_j} - 1], \ldots, j_n),$$

so that

$$\mu(j') = \frac{\mu(j)}{x_{k_j}},$$

or equivalently

$$x_{k_j}\mu(j') = \mu(j).$$

As such, this represents the inverse of the process of building the ranges that make up row $\sigma(j) + 1$.

Continuing one stage further, if j' satisfies

$$\nu(j') = (0, \ldots, 0, j'_{k_{j'}}, \ldots, j'_n), \qquad j'_{k_{j'}} > 0,$$

then we will define j'' by

$$\nu(j'') = (0, \ldots, 0, [j'_{k_{j'}} - 1], \ldots, j'_n)$$

so that

$$\mu(j'') = \frac{\mu(j')}{x_{k'_j}},$$

or equivalently

$$x_{k_{j'}}\mu(j'') = \mu(j').$$

The indices j, k_j, j', $k_{j'}$, and j'' play a defining role in our construction.

To compare with Weisfeld's ordering (1.2):

LEMMA 1. Consider

$$\nu(j^*) = (j_1^*, \ldots, j_n^*) \qquad \text{and} \qquad \nu(j^{**}) = (j_1^{**}, \ldots, j_n^{**}).$$

The ordering in the table is such that

$$\nu(j^*) < \nu(j^{**}) \quad (\text{and } j^* < j^{**} \text{ and } \mu(j^*) < \mu(j^{**}))$$

if and only if either

$$\sigma(j^*) < \sigma(j^{**}) \tag{2.1}$$

or else

$$\sigma(j^*) = \sigma(j^{**}) \tag{2.2a}$$

and

$$j_k^* > j_k^{**}, \tag{2.2b}$$

where k is the *smallest* index such that $j_k^* \neq j_k^{**}$.

This is easily seen by direct verification from Definition 1.

3. INDEXING

Our construction will make use of the indices j, k_j, j', and j'' as given above. But in generating the orthogonal multinomials ψ_j, and in the evaluation of any general multinomial

$$\Psi = \sum_{j=1}^{m} c_j \psi_j,$$

it will not be necessary to access the indices j, k_j, j', and j'' randomly; we may always proceed along the table of monomials from top to bottom and left to right, maintaining and updating these indices as we go.

The k_jth range of row $r = \sigma(j) + 1$, $r > 1$, is constructed to consist of all monomials in which $x_1, \ldots, x_{k_j-1}$ appear to the power 0; that is, it is associated with the exponent vectors $\nu(j)$ of the form $(0, \ldots, 0, j_{k_j}, \ldots, j_n)$ with $j_{k_j} > 0$. In the rth row we must have the equality

$$j_{k_j} + \cdots + j_n = r - 1.$$

Hence the number of items in the k_jth range is equivalent to the number of ways in which $r - 1$ counters (indistinguishable balls) can be distributed among $n - k_j + 1$ exponent positions (distinguishable urns), requiring that at least one counter be assigned to the first position. This number is given by the binomial coefficient

$$\binom{n - k_j + r - 2}{r - 2}.$$

The number of entries in range k_j through n is the number we obtain by removing the restriction that a counter be placed in the first position:

$$\sum_{l=k_j}^{n} \binom{n - l + r - 2}{r - 2} = \binom{n - k_j + r - 1}{r - 1}.$$

And, from this, the number of entries in the entire rth row is given by

$$\binom{n+r-2}{r-1},$$

and the number of entries in all of the first r rows together is

$$\binom{n+r-1}{n}.$$

That is, if $T = n - 1 + r - 2$ and $B = r - 2$, the run lengths are of the form

$$\text{Row } r: \quad \binom{T}{B} \quad \binom{T-1}{B} \quad \binom{T-2}{B} \cdots,$$

$$\text{Row } r+1: \quad \binom{T+1}{B+1} \quad \binom{T}{B+1} \quad \binom{T-1}{B+1} \cdots,$$

and the row lengths are of the form

$$\text{Row } r: \quad \binom{T+1}{B+1},$$

$$\text{Row } r+1: \quad \binom{T+2}{B+2}.$$

Thus, the relationships

$$\binom{T-1}{B} = \binom{T}{B}\frac{T-B}{T}$$

and

$$\binom{T+1}{B+1} = \binom{T}{B}\frac{T+1}{B+1}$$

are clearly useful.

As j runs through range k_j in row r, j' runs in step through ranges $k_j, \ldots, n$ in row $r-1$. As j crosses from range k_j to range $k_j + 1$, j' must be set back to the beginning of range $k_j + 1$ in row $r - 1$. Similarly, as j' runs along row $r-1$, j'' runs in step along row $r-2$, and whenever j' crosses from range $k_{j'}$ to range $k_{j'+1}$ in row $r-1$, then j'' must be reset to the beginning of range $k_{j'+1}$ in row $r-1$.

The code below displays how one can march through the monomial table in order from left to right and from top to bottom.

```
{/
   A program to step through the indices in the ordering table
   for the orthogonal multinomials. The orthogonal multinomials
   are generated according to the pattern

psi[j] = x[kj]*psi[jprime]
         - sum(index t running from jprimeprime to j - 1:
         a[j, t]*psi[t])
      dimen  number of variables
     degree  maximum number of rows to be used in table
```

```
        npolys  size of table
             j  current position in table
            kj  current range in row
        jprime  distinguished multinomial in previous row
   jprimeprime  lower limit on summation
         ralen  current range length in current row
        rowlen  length of current row
           jsw  position of start of next range
        indexs  array for storage of kj, jprime, jprimeprime
           top  T, the top of the binomial coefficient giving
                rowlen
           bot  B, the bottom of this binomial coefficient

  Note: The value of ralen is computed directly from the bino-
    mial coefficient binomial(top, bot), while the value of
    rowlen is obtained by updating this binomial coefficient.

  Note: npolys = binomial(dimen + degree - 1, dimen).

  Note:
    indexs[1, j] will store jprime values;
    indexs[2, j] will store kj values;
    indexs[1, indexs[1, j]] will give the jprimeprime values
     (for j > dimen + 1).

  Note: constants are currently set for the 3-variable, 4-row
    problem.
}

const
  dimen = 3;
  degree = 4;
  npolys = 20;
var
  j, jbeg, jend, jprime, jprimeprime, jsw,
  kj, curdeg, ralen, rowlen, top, bot : integer;
  indexs : array [1..2, 1..npolys] of integer;
begin
  rowlen := 1;
  j := 1;
  jprime := 0;
  top := dimen - 1;
  bot := 0;
  indexs[1, 2] := 1;

  for curdeg := 2 to degree do
  begin
    jbeg := j + 1;
    kj := 1;
    top := top + 1;
    bot := bot + 1;
    ralen := rowlen;
    rowlen := (rowlen*top)div bot;
    jsw := j + ralen;
    jend := j + rowlen - 1;
    for j := jbeg to jend do
    begin
      jprime := jprime + 1;
```

```
      if (j ≥ jsw) then
    begin
      ralen := (ralen*(top - kj - bot + 1)) div (top - kj);
      kj := kj + 1;
      jsw := j + ralen;
      jprime := jprime - ralen
    end;
    indexs[j, 1] := kj;
    indexs[j, 2] := jprime;
    jprimeprime := indexs[jprime, 2];
   end

  end
end.
```

4. THE MULTINOMIAL GRAM–SCHMIDT PROCESS

We propose that the Gram–Schmidt process be

$$\psi_1 = 1,$$

and for $j = 1, 2, \ldots,$

$$\psi_j = x_{k_j}\psi_{j'} - \sum_{l=1}^{j-1} \alpha_{j,l}\psi_l, \tag{4.1}$$

where

$$\alpha_{j,l} = \frac{\langle x_{k_j}\psi_{j'}, \psi_l\rangle}{\langle \psi_l, \psi_l\rangle}$$

and where k_j and j' are related to j according to the indexing program of the preceding section.

To analyze this version of the Gram–Schmidt process, we establish the following lemmas.

LEMMA 2. Let

$$\mu(l) < \mu(m),$$

and let x_p be any of the variables of the problem. Then

$$x_p\mu(l) < x_p\mu(m).$$

PROOF. If

$$\sigma(l) < \sigma(m),$$

then

$$\sigma(\mu^{-1}(x_p)\mu(l)) = \sigma(l) + 1 < \sigma(\mu^{-1}(x_p)\mu(m)) = \sigma(m) + 1.$$

So assume that $\sigma(l) = \sigma(m)$; that is, $\mu(l)$ and $\mu(m)$ are found in the same row of the table. If $\mu(l) < \mu(m)$, then $\nu(l)$ and $\nu(m)$ are such that

$$l_k > m_k \tag{4.2}$$

for the first entry, k, from the left at which these two exponent vectors differ.

But the exponent vectors of $x_p\mu(l)$ and $x_p\mu(m)$ are those of $\mu(l)$ and $\mu(m)$, respectively, with a 1 added into position p. This does not change the role of the index k or the relationship in (4.2) above. □

COROLLARY 1. If

$$\mu(l) < \mu(j'),$$

then

$$x_{k_j}\mu(l) < x_{k_j}\mu(j') = \mu(j),$$

and if

$$\mu(l) < \mu(j''),$$

then

$$x_{k_j}\mu(l) < x_{k_{j'}}\mu(j'') = \mu(j').$$

LEMMA 3. If x_p and x_q are variables of the problem, and $\mu(l)$is any monomial, then

(i) if $p \le q$, then $x_p\mu(l) \le x_q\mu(l)$;
(ii) if $p < q$, then $x_p\mu(l) < x_q\mu(l)$.

PROOF. The exponent vector for $x_p\mu(l)$ would be

$$(l_1, \ldots, l_{p-1}, [l_p + 1], l_{p+1}, \ldots, l_q, \ldots, l_n),$$

whereas that for $x_q\mu(l)$ would be

$$(l_1, \ldots, l_p, \ldots, l_{q-1}, [l_q + 1], l_{q+1}, \ldots, l_n).$$

Hence the first exponent vector comes before the second in the ordering of the table (unless $p = q$, in which case the two exponent vectors are equal). □

COROLLARY 2. If $p \le q$ and $l \le m$, then $x_p\mu(l) \le x_q\mu(m)$. If, in addition, $p < q$ or $l < m$, then $x_p\mu(l) < x_q\mu(m)$.

PROOF. By Lemma 3

$$x_p\mu(l) \le x_q\mu(l).$$

By Lemma 2

$$x_q\mu(l) \le x_q\mu(m).$$

If $p < q$ or $l < m$, this makes at least one of these inequalities sharp. □

LEMMA 4. The Gram–Schmidt process (4.1) yields

$$\psi_j = \mu(j) + \sum_{l=1}^{j-1} \delta_l\mu(l) \tag{4.3}$$

for some coefficients δ_l.

PROOF (by induction).
For $j = 1$:

$$\psi_j \equiv 1 \qquad \text{and} \qquad \mu(j) \equiv 1,$$

and (4.3) holds trivially.

For $j > 1$: Suppose that the result has been established for $1 \le l \le j - 1$. Consider

$$\psi_j = x_{k_j}\psi_{j'} - \sum_{l=1}^{j-1} \alpha_{j,l}\psi_l$$

But $x_{k_j}\mu(j') = \mu(j)$. Furthermore, $j' \le j - 1$, so by hypothesis

$$\psi_{j'} = \mu(j') + \sum_{l=1}^{j'-1} \delta_l' \mu(l).$$

Hence,

$$\psi_j = \mu(j) + \sum_{l=1}^{j'-1} \delta_{l'} x_{k_j}\mu(l) - \sum_{l=1}^{j-1} \alpha_{j,l}\psi_l.$$

But from Corollary 1

$$x_{k_j}\mu(l) < \mu(j)$$

for all $1 \le l \le j' - 1$, and, by the induction hypothesis, each ψ_l in the right-hand summation can be expressed as a linear combination of the 1st through $(j - 1)$th monomials. The result follows by collecting terms in the individual monomials. □

LEMMA 5. The multinomials ψ_j are linearly independent.

This is evident from Lemma 4.

From the foregoing we can establish

LEMMA 6. If $x_{k_j}\mu(l) < \mu(j')$, then $\alpha_{j,l} = 0$.

PROOF. $\alpha_{j,l} = 0$ means $\langle x_k\psi_{j'}, \psi_l\rangle = 0$.
But

$$\langle x_{k_j}\psi_{j'}, \psi_l\rangle = \langle \psi_{j'}, x_{k_j}\psi_l\rangle.$$

By Lemma 2,

$$x_{k_j}\mu(1) < \cdots < x_{k_j}\mu(l), \qquad \text{and} \qquad x_{k_j}\mu(l) < x_{k_j}\mu(j'').$$

Hence,

$$x_{k_j}\mu(l) \le x_{k_j}\mu(j' - 1).$$

So $x_{k_j}\psi_l$ will be in the span of $\{\mu(1), \ldots, \mu(j' - 1)\} \equiv \{\psi_1, \ldots, \psi_{j'-1}\}$. □

THEOREM 1. For all $l < j''$,

$$\alpha_{j,l} = 0.$$

PROOF. $l < j''$ and $k_j \leq k_{j'}$. So, by Corollary 2, $x_{k_j}\mu(l) < \mu(j')$, and the result follows from Lemma 6. □

Some results from data with $n = 2$ are given in tabular outline below, where "*" denotes the position of the orthogonal multinomial being constructed, "+" denotes the position of a previous orthogonal multinomial whose associated α is computed to be nonzero, and "0" denotes the position of a previous orthogonal multinomial whose associated α is computed as zero. The position of j'' is indicated by "!".

```
j = 1:  *
j = 2:  +          j = 3:  +
        *                  + *
j = 4:  +!         j = 5:  0!
        + +                + +
        *                  + *
j = 6:  +!         j = 7:  0
        + +                +! +
        + + *              + + +
                           *
j = 8:  0          j = 9:  0
        0 +!               0 0!
        + + +              + + +
        + *                + + *
j = 10: 0          j = 11: 0
        0 +!               0 0
        + + +              +! + +
        + + + *            + + + +
                           *
```

It is visible from this that certain extra α's will turn out to be zero. What is not visible in the above schema is that many of the nonzero α's have related values, which could be used to save storage and computation. The next section will establish some results.

5. ADDITIONAL RELATIONSHIPS

We made use of the fact that $k_j \leq k_{j'}$ in establishing Theorem 1 of the preceding section. We can sharpen the result of that theorem by splitting this inequality up into its two possible cases.

COROLLARY 3. Assume that

$$k_j = k_{j'} \quad \text{and} \quad \sigma(j) \geq 2.$$

Then

$$\alpha_{j,j''} \neq 0.$$

PROOF. Consider

$$\psi_{j'} = x_{k_{j'}}\psi_{j''} - \sum_{l=1}^{j'-1} \alpha_{j',l}\psi_l.$$

It follows from this and the orthogonality of the ψ's that

$$\begin{aligned}\langle \psi_{j'}, \psi_{j'} \rangle &= \langle x_{k_{j'}}\psi_{j''}, \psi_{j'} \rangle \\ &= \langle x_{k_{j'}}\psi_{j'}, \psi_{j''} \rangle \\ &= \langle x_{k_j}\psi_{j'}, \psi_{j''} \rangle \\ &= \alpha_{j,j''}\langle \psi_{j''}, \psi_{j''} \rangle.\end{aligned}$$

Since both inner products are positive,

$$\alpha_{j,j''} = \frac{\langle \psi_{j'}, \psi_{j'} \rangle}{\langle \psi_{j''}, \psi_{j''} \rangle} \neq 0. \qquad \square$$

COROLLARY 4. Assume that

$$k_j < k_{j'}.$$

Then

$$\alpha_{j,l} = 0$$

for l running from j'' out to the end of the row containing j'', that is, for

$$\mu(j'') \leq \mu(l) \leq x_n^{\sigma(j)-2}.$$

PROOF. In this case we have

$$\begin{aligned}&\mu(j) \quad \text{associated with} \quad (0, \ldots, 0, [1], 0, \ldots, 0, [j_{k_{j'}}], \ldots, j_n), \\ &\mu(j') \quad \text{associated with} \quad (0, \ldots, 0, [0], 0, \ldots, 0, [j_{k_{j'}}], \ldots, j_n), \\ &\mu(j'') \quad \text{associated with} \quad (0, \ldots, 0, [0], 0, \ldots, 0, [j_{k_{j'}} - 1], \ldots, j_n),\end{aligned}$$

where [1] and [0] mark the k_jth component.

But $x_{k_j} < x_{k_{j'}}$ and $\mu(l) \leq \mu(j'')$. So $x_{k_j}\mu(l) < \mu(j')$ by Corollary 2, and $\alpha_{j,l} = 0$ by Lemma 6. $\square$

Notice that, if indices j, l, p, and m are such that

$$\langle x_{k_j}\psi_{j'}, \psi_l \rangle = \langle x_{k_p}\psi_{p'}, \psi_m \rangle \tag{5.1}$$

then

$$\alpha_{j,l} = \alpha_{p,m} \frac{\langle \psi_m, \psi_m \rangle}{\langle \psi_l, \psi_l \rangle}.$$

To explore such associations, we have generated numbers of examples with random data. To give an instance, using three variables up through the first four

rows of the table yielded the following associations:

$$\alpha_{6,2} \leftrightarrow \alpha_{5,3} \qquad \alpha_{15,5} \leftrightarrow \alpha_{11,9}$$

$$\alpha_{7,2} \leftrightarrow \alpha_{5,4} \qquad \alpha_{15,6} \leftrightarrow \alpha_{12,9}$$

$$\alpha_{7,3} \leftrightarrow \alpha_{6,4} \qquad \alpha_{15,7} \leftrightarrow \alpha_{13,9}$$

$$\alpha_{9,3} \leftrightarrow \alpha_{8,4} \qquad \alpha_{15,8} \leftrightarrow \alpha_{14,9}$$

$$\alpha_{11,3} \leftrightarrow \alpha_{6,5} \qquad \alpha_{16,5} \leftrightarrow \alpha_{11,10}$$

$$\alpha_{11,4} \leftrightarrow \alpha_{7,5} \qquad \alpha_{16,6} \leftrightarrow \alpha_{12,10}$$

$$\alpha_{12,4} \leftrightarrow \alpha_{7,6} \qquad \alpha_{16,7} \leftrightarrow \alpha_{13,10}$$

$$\alpha_{12,5} \leftrightarrow \alpha_{11,6} \qquad \alpha_{16,8} \leftrightarrow \alpha_{14,10}$$

$$\alpha_{13,5} \leftrightarrow \alpha_{11,7} \qquad \alpha_{16,9} \leftrightarrow \alpha_{15,10}$$

$$\alpha_{13,6} \leftrightarrow \alpha_{12,7} \qquad \alpha_{17,4} \leftrightarrow \alpha_{9,8}$$

$$\alpha_{14,5} \leftrightarrow \alpha_{11,8} \qquad \alpha_{18,8} \leftrightarrow \alpha_{17,9}$$

$$\alpha_{14,6} \leftrightarrow \alpha_{12,8} \qquad \alpha_{19,8} \leftrightarrow \alpha_{17,10}$$

$$\alpha_{14,7} \leftrightarrow \alpha_{13,8} \qquad \alpha_{19,9} \leftrightarrow \alpha_{18,10}$$

In each of the above the association is such that

$$\alpha_{j,l} \leftrightarrow \alpha_{p,m}$$

if and only if

$$m = j', \qquad l = p' \tag{5.2}$$

and

$$k_j = k_p.$$

It is trivial to show that the α's associate when j, l, and p satisfy these requirements. Apparently no other α's associate.

This association will be of use primarily in telling us that the inner product

$$\langle x_{k_j} \psi_{j'}, \psi_l \rangle$$

has been computed before.

As a first consideration, the case

$$\mu(l) < \mu(j'')$$

can be ignored, since

$$\langle x_{k_j} \psi_{j'}, \psi_l \rangle = 0$$

for all such l. Second, the case

$$\mu(l) = \mu(j'')$$

is not of interest, since it is handled fully in Corollary 2 and Corollary 3. Third, the case

$$\mu(j'') \leq \mu(l) \leq x_n^{\sigma(j)-2}$$

is not of interest when $k_j < k_{j'}$, since

$$\langle x_{k_j}\psi_{j'}, \psi_l\rangle = 0$$

by Corollary 3. As a fourth consideration, (5.2) will require that

$$\mu(l) \geq x_{k_j}^{\sigma(j)-1}$$

when $\mu(l)$ falls in the row of j'. And finally, we must have

$$\mu(p) = x_{k_j}\mu(l) < \mu(j),$$

since we are only interested in inner products that have been computed in the past and can be reused. This implies that

$$\mu(l) < \mu(j').$$

Hence we are left only with the cases

(A) $x_{k_j}^{\sigma(j)-2} < \mu(l) \leq x_n^{\sigma(j)-2}$ and $k_j = k_{j'}$;

(B) $x_{k_j}^{\sigma(j)-1} \leq \mu(l) < \mu(j')$.

The examples of $\alpha_{16,l}$ are illustrative. Since

$$\mu(16) = x_1\mu(10)$$

and

$$\mu(10) = x_3\mu(4),$$

we have

$$j = 16, \quad j' = 10, \quad k_j = 1, \quad j'' = 4, \quad k_{j'} = 3.$$

Hence we must have

$$\alpha_{16,l} = \alpha_{p,10}$$

for each $\mu(p)$ that can be expressed as $x_1\mu(l)$. These are precisely those for which (B) above holds:

$$x_1^2 \leq \mu(l) < \mu(10),$$

which yields

$$p = 11, 12, 13, 14, 15$$

for

$$l = 5, 6, 7, 8, 9,$$

respectively.

6. FORTRAN PROGRAM

A program has been prepared to implement the results described in this paper (see Algorithm 634, pages 218–228). All of the features of indexing, attention to zero inner products, and attention to associations between α's have been included as described. In addition, the computation of the fitting coefficients

$$c_l = \frac{\langle F, \psi_l \rangle}{\langle \psi_l, \psi_l \rangle}$$

is done by the alternative formula

$$c_l = \frac{\langle z_{l-1}, \psi_l \rangle}{\langle \psi_l, \psi_l \rangle},$$

where

$$z_{l-1} = F - \sum_{i=1}^{l-1} c_i \psi_i$$

as is advocated in [1]. The program has been passed by the PFORT [5] verifier.

7. BASIS SELECTION

In the case that the function

$$F(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) + e(x_1, \ldots, x_n)$$

that is being fit satisfies

(1) the errors at each point in the sample space $(x_1, \ldots, x_n)$ are $e(x_1, \ldots, x_n) \sim N(0, \sigma^2)$,
(2) for all points in the sample space $(x_1, \ldots, x_n)$, the $e(x_1, \ldots, x_n)$ are independent with the same variance,
(3) $f(x_1, \ldots, x_n)$ is, in fact, a multinomial,

and in the case that the first m orthogonal multinomials generated by our process yield a satisfactory model for the data, that is, f is in their span, then the person interested in doing the least-squares fitting can test the hypothesis that the subset

$$\psi_{i_1}, \psi_{i_2}, \ldots, \psi_{i_r}$$

are the only basis elements necessary to represent f, if

$$\frac{N-m}{r} \frac{\langle F - P_r, F - P_r \rangle - \langle F - P_m, F - P_m \rangle}{\langle F - P_m, F - P_m \rangle} \sim \mathbf{F}_{r,N-m}$$

is true, where $\mathbf{F}$ is the F-distribution, P_r is the least-squares fitting multinomial produced using only the ψ_{i_j}, and P_m is the least-squares fitting multinomial produced from the full basis $\psi_1, \ldots, \psi_m$.

The above accords with the description that Forsythe gives in [3] about selecting the order of the fitting polynomial; Forsythe's description treats the special case where the basis subset consists of $\psi_1, \ldots, \psi_r$. A reference for F-testing is to be found on [2, pp. 307–310]. (For a routine that was written to administer the selection of subsets in regression problems according to the

material in [2], refer to the routine RLEAP in the IMSL [4] and to the further bibliographic references given for that routine.)

The FORTRAN code referred to in the previous section was written to be efficient for a limited brand of incremental fitting. It is possible to obtain the least-squares fit using the subset $\psi_1, \ldots, \psi_{r+1}$ from the fit obtained using the subset $\psi_1, \ldots, \psi_r$ with a minimum of computational effort.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Colin DiCenzo, formerly of the Canadian Government Atmospheric Environment Service, for suggesting the value of studying this data-fitting problem and for pointing out the work of Weisfeld. The authors have also had helpful discussions with J. G. Kalbfleisch of the University of Waterloo Statistics Department.

REFERENCES

1. CADWELL, J. H., AND WILLIAMS, D. E. Some orthogonal methods of curve and surface fitting. *Comput. J. 4* (1961), 260–261.
2. DRAPER, N. R., AND SMITH, H. *Applied Regression Analysis.* Wiley, New York, 1980.
3. FORSYTHE, G. E. Generation and use of orthogonal polynomials for data-fitting with a digital computer. *J. SIAM 5* (1957), 74–87.
4. IMSL. *RLEAP.* The International Mathematical and Statistical Libraries, Houston, Texas, 1984.
5. RYDER, B. G. The PFORT verifier. *Softw. Pract. Exper. 4* (1974), 359–377.
6. WEISFELD, M. Orthogonal polynomials in several variables. *Numer. Math. 1* (1959), 38–40.

Received April, 1984; revised December 1984; accepted February 1985