Gauss elimination by segments and multivariate polynomial interpolation

C. de Boor [1]

Abstract. The construction of a polynomial interpolant to data given at finite pointsets in $\mathbb{R}^d$ (or, most generally, to data specified by finitely many linear functionals) is considered, with special emphasis on the linear system to be solved. Gauss elimination by segments (i.e., by groups of columns rather than by columns) is proposed as a reasonable means for obtaining a description of all solutions and for seeking out solutions with 'good' properties. A particular scheme, due to Amos Ron and the author, for choosing a particular polynomial interpolating space in dependence on the given data points, is seen to be singled out by requirements of degree-reduction, dilation-invariance, and a certain orthogonality requirement. The close connection, between this particular construction of a polynomial interpolant and the construction of an H-basis for the ideal of all polynomials which vanish at the given data points, is also discussed.

1. Introduction

Polynomial interpolation in d variables has long been studied from the following point of view: one is given a polynomial space, F, and seeks to characterize the point sets $\Theta \subset \mathbb{R}^d$ for which the pair (Θ, F) is **correct** in the sense that F contains, for each g defined at least on Θ, exactly one element which agrees with g on Θ. Further, the space F is either the space Π_k of all polynomials on $\mathbb{R}^d$ of degree $\le k$, or is, more generally, a D-invariant space spanned by monomials. See R. A. Lorentz [9] for an up-to-date accounting of the many interesting efforts in this direction.

In contrast, de Boor and Ron [5] starts with an arbitrary finite point set $\Theta \subset \mathbb{R}^d$ and proposes a particular choice, denoted Π_Θ, from among the many polynomial spaces F for which (Θ, F) is correct. To be explicit (an *explanation* can be found at the end of Section 7), the definition of Π_Θ makes use of the **least** term $f_\downarrow$ of a function f, which, by definition, is the first, or lowest-degree, nontrivial term in the expansion of f into a sum $f = f_0 + f_1 + f_2 + \ldots$ in which f_j is a homogeneous polynomial of degree j, all j. With this, Π_Θ is defined as the linear span

$$\Pi_\Theta := \operatorname{span}\{f_\downarrow : f \in \exp_\Theta\}$$

of least terms of elements in the linear span $\exp_\Theta$ of exponentials $e_\theta : x \mapsto \exp(\theta \cdot x)$, $\theta \in \Theta$. Because of this, we have come to call the resulting interpolant

$$P_\Theta g$$

[1] Computer Sciences Department, 1210 W. Dayton St., Madison WI 53706-1685, USA (deboor@cs.wisc.edu)

to a given function g, defined at least on Θ, the **least interpolant to** g.

The connection between the least interpolant and Gauss elimination (applied to the appropriate Vandermonde matrix) was already explored in de Boor [2] and de Boor and Ron [7]. However, this earlier discussion left the (wrong) impression that a particular variant of Gauss elimination was needed in order to obtain the least interpolant. The present paper corrects this impression, in the process of examining the question of what might single out Π_Θ from among all polynomial spaces F for which (Θ, F) is correct. To be sure, de Boor and Ron [8] already contains such a discussion, but not at all from the point of view of elimination. As in [8], the discussion here is actually carried out in the most general situation possible, namely of interpolation to data of the form Λg, with Λ an arbitrary linear map to $\mathbb{R}^n$ from the space Π of all polynomials on $\mathbb{R}^d$. Since such 'data maps' are of finite rank, it would be trivial to extend them to functions on $\mathbb{R}^d$ other than polynomials.

In the discussion, two properties of Π_Θ are seen to play a particularly important role: (i) P_Θ is **degree-reducing**, i.e.,

$$\deg P_\Theta g \le \deg g, \quad \forall\, g \in \Pi;$$

and (ii) Π_Θ is **dilation-invariant**, i.e., for any $r > 0$ and any $p \in \Pi_\Theta$, also $p(r\cdot) \in \Pi_\Theta$. The latter property is equivalent to the fact that Π_Θ **stratifies**, i.e., Π_Θ is the direct sum $\Pi_\Theta = \bigoplus_k \Pi^0_{\Theta,k}$ of its homogeneous subspaces

$$\Pi^0_{\Theta,k} := \Pi_\Theta \cap \Pi^0_k$$

(with Π^0_k the space of all *homogeneous* polynomials of exact degree k (including the zero polynomial)). Also, the most intriguing property of Π_Θ (see de Boor and Ron [8]), namely that

$$\Pi_\Theta = \bigcap_{p|_\Theta = 0} \ker p_\uparrow(D),$$

is looked at anew. Here, $p_\uparrow$ is the **leading** term of the polynomial p, i.e., the last, or highest-degree, nontrivial term in the expansion $p = p_0 + p_1 + p_2 + \ldots$ of p with p_j homogeneous of degree j, all j.

In addition, it is shown how to make use of the construction of Π_Θ by elimination to construct an H-basis for the polynomial ideal of all polynomials which vanish on Θ. This responds to questions raised by algebraists during talks on the algorithm given in [2] and [7].

The paper highlights *Gauss elimination by segments.* In this generalization of Gauss elimination, the matrix A to be factored is somehow **segmented**, i.e., is given in the form

$$A = [A_0, A_1, \ldots],$$

with each **segment**, A_j, comprising zero or more consecutive columns of A. (Here and throughout, we use MATLAB notation.) Correspondingly, elimination is to proceed, not column by column, but segment by segment. In the case of polynomial interpolation, the segmentation naturally arises by grouping monomials of the same degree together and because there is no natural ordering for the monomials of the same degree. Because of this particular application, we start the indexing of the segments of A with 0.

By definition, **Gauss elimination by segments** applied to $A = [A_0, A_1, \ldots]$ produces a factorization

$$[A_0, A_1, \ldots,] = M[R_0, R_1, \ldots]$$

with M invertible and

$$R := [R_0, R_1, \ldots]$$

a **segmented row-echelon form**, the segmentation corresponding to that of A. This means that R is block upper triangular, $R = (R_{ij})$ say with $R_j = [R_{0j}; R_{1j}; \ldots]$ having exactly as many columns as A_j does, with each diagonal block R_{jj} onto. (A casual search of the literature failed to turn up earlier occurrences of this concept of a segmented row echelon form. However, in hindsight, the notion is so natural that it is bound to have been used before.)

Except for trivial cases, a segmented matrix has infinitely many such factorizations. Nevertheless, all such factorizations have certain properties in common. For our purposes, the most important property is the fact that, for each j, the row-space of R_{jj} depends only on A (and the segmentation). In particular, it is independent of the details of the elimination process which led to R. We take the trouble to prove this since, as we also show, the row-space of R_{jj} provides all the information needed to construct the space $\Pi^0_{\Theta,j} = \Pi_\Theta \cap \Pi^0_j$ of homogeneous polynomials of degree j in Π_Θ. We also identify the numerical procedure proposed in [2] and [7] as a particularly stable way for obtaining such a segmented row-echelon form.

For simplicity, the paper deals with *real*-valued functions only. All the results are true if matrix transpose is replaced by conjugate transpose and, correspondingly, certain quantities by their complex conjugate.

Finally, the paper is also meant to illustrate the perhaps debatable point that it is possible to write about polynomial interpolation in several variables without covering entire pages with formulæ.

The paper had its start in a discussion I had with Nira Dyn, in August of 1992, concerning possible alternatives to the algorithm in [7], and subsequent discussions, in February of 1993, concerning possible alternatives to the least interpolant. In particular, the use of column maps below is my response to Nira Dyn's very direct and useful way of describing the least interpolant.

2. Polynomials

A polynomial in d variables is customarily written in the form

$$p(x) = \sum_{\alpha} x^{\alpha} c(\alpha), \tag{2.1}$$

with

$$x^{\alpha} := x(1)^{\alpha(1)} \cdots x(d)^{\alpha(d)}, \quad x = (x(1), \ldots, x(d)) \in \mathbb{R}^d$$

and the sum taken over the set

$$\mathbb{Z}_+^d = \{\alpha \in \mathbb{R}^d : \alpha(j) \in \mathbb{Z}_+, \text{ all } j\}$$

of multi-indices. Further, the coefficient sequence c has finite support.

Since it is important to distinguish between the polynomial p and its value $p(x)$ at the point x, a notation for the power map $x \mapsto x^{\alpha}$ is needed. There being no standard notation around, we'll use $()^{\alpha}$ for it. Here is the formal definition:

$$()^{\alpha} : \mathbb{R}^d \to \mathbb{R} : x \mapsto x^{\alpha}.$$

In these terms, the polynomial p of (2.1) can be written

$$p = \sum_{\alpha} ()^{\alpha}\, c(\alpha).$$

As indicated already by the unusual positioning of the coefficients, this formula for p can be viewed as the result of applying the 'matrix'

$$X := [()^{\alpha} : \alpha \in \mathbb{Z}_+^d]$$

with 'columns' $()^{\alpha}$ to the 'vector' c in the standard way, namely by multiplying the 'column' with index α with the entry of c with index α and summing. Thus,

$$p = Xc.$$

This notation (taken from de Boor [3]) turns out to be very convenient in what is to follow.

Formally, we think of X as a linear map, defined on the space

$$\operatorname{dom} X := (\mathbb{Z}_+^d \to \mathbb{R})_0 := \{c : \mathbb{Z}_+^d \to \mathbb{R} : \#\operatorname{supp} c < \infty\}$$

of finitely supported real-valued sequences indexed by $\mathbb{Z}_+^d$, and mapping into the space $\mathbb{R}^d \to \mathbb{R}$ of real-valued functions of d arguments.

Note that X is 1-1 and its range is Π, the space of polynomials in d indeterminates, interpreted here as functions on $\mathbb{R}^d$. Any linear subspace F of Π of dimension n is the range of a map XW, with W a 1-1 linear map from $\mathbb{R}^n$ into $\operatorname{dom} X$, hence $V := XW$ (or, in more traditional language, the sequence of columns of V) is a basis for F. To be sure, both W and V are linear maps on $\mathbb{R}^n$, hence are, in the language of [3], **column map**s. This language is meant to stress the fact that any linear map B, from n-dimensional coordinate space $\mathbb{F}^n$ to some linear space S over the scalar-field $\mathbb{F}$, is necessarily of the form

$$B = [b_1, \ldots, b_n] : \mathbb{F}^n \to S : c \mapsto \sum_j b_j c(j),$$

with

$$b_j := B\mathbf{i}_j$$

the j**th column of** B and

$$\mathbf{i}_j := (\underbrace{0, \ldots, 0}_{j-1 \text{ zeros}}, 1, 0, \ldots).$$

We denote by

$$\#B$$

the number of columns in the column map B. If the target, S, of the column map $B \in L(\mathbb{F}^n, S)$ is itself a coordinate space, $S = \mathbb{F}^m$ say, hence B is (naturally identified with) an $m \times n$-matrix, then the columns of B as a column map are, indeed, its columns as a matrix.

For any linear map A, whether a column map or not, we denote by

$$\operatorname{ran} A, \quad \ker A$$

its **range** (or set of values) respectively its **kernel** or nullspace.

3. Interpolation

As discussed in the Introduction, we are interested in interpolation at a given finite point set $\Theta \subset \mathbb{R}^d$, of cardinality

$$n := \#\Theta,$$

say. Define

$$\Lambda : \Pi \to \mathbb{R}^n : g \mapsto g|_\Theta := (g(\theta) : \theta \in \Theta).$$

Then, $f \in \Pi$ interpolates to g at Θ iff f solves the equation

$$\Lambda ? = \Lambda g. \tag{3.1}$$

We have introduced the map Λ here, not only for notational convenience, but also in order to stress the fact that most of the discussion to follow is valid for any onto linear map $\Lambda : \Pi \to \mathbb{R}^n$. We call any linear map $\Lambda : \Pi \to \mathbb{R}^n$ a **data map** and note that it is necessarily of the form $\Lambda : g \mapsto (\lambda_i g : i = 1, \ldots, n)$ for some linear functionals λ_i on Π, which we call the **rows** of Λ, for obvious reasons. Given an onto data map Λ, we say that (Λ, F) is **correct** (or, F **is correct for** Λ) if (3.1) has exactly one solution in F for every $g \in \Pi$. In that case, we denote the unique interpolant in F to $g \in \Pi$ by

$$P_F g.$$

Here, for completeness, are certain known facts about interpolation.

Lemma 3.2. *Let F be an n-dimensional polynomial space, let $V \in L(\mathbb{R}^n, \Pi)$ be a basis for F, and let $\Lambda \in L(\Pi, \mathbb{R}^n)$ be onto. Then, the following are equivalent:*
(i) *(Λ, F) is correct.*
(ii) *$\Lambda|_F$ is invertible.*
(iii) *Λ is 1-1 on F.*
(iv) *ΛV is invertible.*

Further, if any one of these conditions holds, then

$$P_F = V(\Lambda V)^{-1}\Lambda.$$

Proof. Since Λ is onto, i.e., its range is all of $\mathbb{R}^n$, correctness of (Λ, F) is equivalent to having the linear map $F \to \mathbb{R}^n : f \mapsto \Lambda f$ be invertible, and, since both domain and target of this map have the same dimension, n, this is equivalent to having this map be 1-1. Finally, since V is a basis for F, the linear map $V|^F : \mathbb{R}^n \to F : a \mapsto Va$ is invertible. Since $\Lambda V = \Lambda|_F V|^F$, it follows that $\Lambda|_F$ is invertible iff ΛV is invertible. □

In order to investigate what spaces F are correct for Λ, we look at all possible polynomial interpolants to a given g, i.e., at all possible solutions of (3.1). Since X is 1-1, the solutions of (3.1) are in 1-1 correspondence with solutions of the equation

$$\Lambda X? = \Lambda g.$$

This equation is a system of n linear algebraic equations, albeit in infinitely many unknowns, namely the power coefficients $c(\alpha)$, $\alpha \in \mathbb{Z}_+^d$, of the interpolant. Nevertheless, we may apply Gauss elimination to determine all solutions.

4. Gauss elimination

Gauss elimination (with row pivoting) produces a factorization

$$\Lambda X = MR,$$

with M invertible, and R 'in row-echelon form', i.e., 'right-triangular' in the following sense. Each row other than the first is either entirely zero, or else its left-most nonzero entry is strictly to the right of the left-most nonzero entry in the preceding row. The left-most nonzero entry (if any) in a row is called the **pivot element** for that row, and the row is called the **pivot row for** the unknown associated with the column in which the pivot element occurs. An unknown is called **bound** or **free** depending on whether or not it has a pivot row. It is a standard result that an unknown is bound if and only if its column is not in the linear span of the columns to the left of it.

Since ΛX is onto (X being onto Π and Λ being onto), every row of R will be a pivot row. Let $(\alpha_1, \ldots, \alpha_n)$ be the corresponding sequence of indices of bound unknowns, and set $F = \operatorname{ran} V$ with

$$V := [()^{\alpha_1}, \ldots, ()^{\alpha_n}].$$

Then ΛV is invertible, hence (Λ, F) is correct.

This construction is quite arbitrary. As should have been clear from the language used, Gauss elimination depends crucially on the *ordering* of the columns of ΛX, i.e., on the ordering of the columns of X. On the other hand, for $d > 1$, there is no natural ordering of the columns of X, i.e., of the monomials. The best we can do, offhand, is to order the monomials by degree. Here, to be sure,

$$\deg Xc = \max\{|\alpha| := \|\alpha\|_1 = \sum_i \alpha(i) : c(\alpha) \neq 0\}.$$

Equivalently, $\deg p = \min\{k : p \in \Pi_k\}$, with

$$\Pi_k := \operatorname{ran} X_{\leq k},$$

where

$$X_{\leq k} := [X_0, \ldots, X_k]$$

and

$$X_j := [()^\alpha : |\alpha| = j].$$

Assume that X is **graded**, i.e., segmented by degree, i.e.,

$$X = [X_0, X_1, \ldots].$$

This has the following happy consequence.

Result 4.1 **([7; (2.5)]).** *If X is graded, then polynomial interpolation from the space F spanned by the monomials corresponding to the bound columns of ΛX is degree-reducing, i.e.,*

$$\deg P_F g \le \deg g \qquad \forall g \in \Pi.$$

Proof. Recall, as we did earlier, that an unknown is bound if and only if its column is not a linear combination of preceding columns. In particular, any column is in the linear span of the bound columns not to the right of it. This implies that, if $\deg g = k$ and therefore $\Lambda g \in \operatorname{ran} \Lambda X_{\le k}$, then Λg is necessarily in the linear span of the bound columns in $\Lambda X_{\le k}$. Since $f := P_F g$ is the unique element in F with $\Lambda f = \Lambda g$, it follows that $P_F g$ is already in

$$F_k := F \cap \Pi_k.$$

□

In particular, for $d = 1$ and $\Lambda = |_\Theta$, we recover in this way the standard choice for F, namely $\Pi_{<n}$. However, for $d > 1$, F still depends on the ordering of the columns within each X_k and there is no natural ordering for them. It is for this reason that we now consider Gauss elimination by segments.

5. Gauss elimination by segments

To recall from the Introduction, Gauss elimination by segments applied to the segmented matrix $A = [A_0, A_1, \ldots]$ produces a factorization

$$[A_0, A_1, \ldots] = M[R_0, R_1, \ldots]$$

with M invertible and

$$R := [R_0, R_1, \ldots]$$

a segmented row-echelon form, the segmentation corresponding to that of A. This means that R is block upper triangular, $R = (R_{ij})$, say, with

$$R_j = [R_{0j}; R_{1j}; \ldots] = [U_{jj}; R_{jj}; 0] := \begin{bmatrix} U_{jj} \\ R_{jj} \\ 0 \end{bmatrix},$$

$\#R_j = \#A_j$, and with each diagonal block R_{jj} onto. (The semicolon is used here as in MATLAB.)

While, except for trivial cases, a segmented matrix has infinitely many such factorizations, some properties of such a factorization depend only on A and the particular segmentation used. Here are the basic facts.

Lemma 5.1. *Let A, B, C, M, S, Q, T be matrices with*

$$A = [B, C] = M \begin{bmatrix} S & Q \\ 0 & T \end{bmatrix},$$

M invertible, $\#B = \#S$, and S onto. Then

(i) The transpose S' of S is (i.e., the rows of S form) a basis for the row-space of B, i.e., for $\operatorname{ran} B'$.

(ii) The row-space of T, i.e., the space $\operatorname{ran} T'$, depends only on A and $\#T = \#A - \#B$. Explicitly,

$$\operatorname{ran} T' = \{y \in \mathbb{R}^{\#T} : (0, y) \in \operatorname{ran} A'\}.$$

(iii) For all $c \perp \operatorname{ran} T'$, there exists b so that $(b, c) \in \ker A$.

Proof. (i): Since $\#B = \#S$, we have $B = MS$, with M invertible, hence $\operatorname{ran} B' = \operatorname{ran} S'M' = \operatorname{ran} S'$.

(ii): Since M is invertible, $\operatorname{ran} A' = \operatorname{ran} \begin{bmatrix} S' & 0 \\ Q' & T' \end{bmatrix}$. In particular, for any $\#T$-sequence y and any $\#S$-sequence x, $(x, y) \in \operatorname{ran} A'$ if and only if

$$(x, y) = (S'z_1, Q'z_1 + T'z_2)$$

for some sequence $z = (z_1, z_2)$. In particular, since S is onto and therefore S' is 1-1, such x is zero iff $z_1 = 0$. In other words, $(0, y) \in \operatorname{ran} A'$ if and only if $y = T'z_2$ for some z_2.

(iii): If $c \perp \operatorname{ran} T'$ then $c \in \ker T$, therefore, for any $\#S$-vector b, $A(b, c) = M(Sb + Qc, 0)$. It follows that $(b, c) \in \ker A$ if and only if $Sb = -Qc$ and, since S is onto by assumption, such a choice for b is always possible. □

Corollary 5.2. *Let*

$$\Lambda X = MR$$

with M invertible and $R = (R_{ij})$ a segmented row-echelon form, segmented corresponding to the segmentation $X = [X_0, X_1, \ldots]$ of X. Then,

(i) for any k, R'_{kk} is a basis for

$$\{y \in \mathbb{R}^{r_k} : (0, y) \in \operatorname{ran}(\Lambda X_{\leq k})'\},$$

with $r_k := \#R_{kk}$. In particular, $\operatorname{ran} R'_{kk}$ depends only on Λ.

(ii) for any k, $\#R'_{kk} = \operatorname{rank} R_{kk} = \operatorname{rank} \Lambda X_{\leq k} - \operatorname{rank} \Lambda X_{<k}$.

Proof. (i): This follows from (ii) of Lemma 5.1, with the choices $B := \Lambda X_{<k}$, $C := \Lambda X_k$, $S := (R_{ij} : 0 \leq i, j < k)$, $Q := [R_{0k}; \ldots; R_{k-1,k}]$, and $T := R_{kk}$.

(ii): Since M is invertible,

$$\operatorname{rank} \Lambda X_{\leq k} = \operatorname{rank} R_{\leq k},$$

and the latter number equals $\sum_{j\le k}\#R'_{jj}$ since $R_{\le k} = (R_{ij} : i = 0,1,\ldots; j = 0,\ldots,k)$ is block upper triangular with each diagonal block onto. In particular, each R_{jj} has full row rank, hence $\#R'_{jj} = \operatorname{rank} R_{jj}$, all j. □

We briefly discuss the construction of a segmented row-echelon form for a given segmented matrix $A = [A_0, A_1, \ldots]$.

It follows from Lemma 5.1 that, in step 0 of Gauss elimination by segments applied to the segmented matrix $A = [A_0, A_1, \ldots]$, one determines, by some means, some basis R'_{00} for $\operatorname{ran} A'_0$. While this could, of course, be done by Gauss elimination, with the nontrivial rows of the resulting row-echelon-form for A_0 providing the desired basis, i.e., providing the rows for the matrix R_{00}, more stable procedures come to mind. For example, one could construct R'_{00} as an orthogonal basis, or even an orthonormal basis, for the row-space of A_0. The favored numerical procedure for this would be to construct the QR factorization for A'_0, preferably using Householder reflections and column interchanges, and use the resulting orthonormal basis for $\operatorname{ran} A'_0$.

Since R'_{00} is a basis for $\operatorname{ran} A'_0$, there is a unique N_0 so that $A_0 = N_0 R_{00}$. In fact, N_0, or some left inverse for N_0, is usually found during the construction of R_{00}. In any case, since the rows of R_{00} are a basis for the row-space of A_0, N_0 is necessarily 1-1, hence can be extended to some invertible matrix $M_0 = [N_0, \ldots]$. This gives the factorization

$$A = [A_0, A_1, A_2, \cdots] = M_0 \begin{bmatrix} R_{00} & & Q & \\ 0 & T_1 & T_2 & \cdots \end{bmatrix}.$$

Subsequent steps concentrate on the segmented matrix $T := [T_1, T_2, \cdots]$.

This discussion of Gauss elimination by segments was only given in order to indicate the many different ways available for constructing a factorization $A = MR$ of a segmented matrix $A = [A_0, A_1, \ldots]$ into an invertible M and a correspondingly segmented row-echelon form. The *existence* of such a factorization was never in doubt, as any factorization $A = MR$ obtained by Gauss elimination with partial pivoting is seen to be of the desired form after R is appropriately blocked.

6. A "natural" choice for F

We have already seen that a correct F can be so chosen that P_F is degree-reducing. The following result characterizes all such choices F.

Result 6.1 ([8]). *If (Λ, F) is correct, then, P_F is degree-reducing if and only if each $F_k = F \cap \Pi_k$ is as large as possible.*

We give a simple proof of this in a moment, for completeness. But first we explore what limits if any might actually be imposed on the dimension of F_k. Since correctness of (Λ, F) implies that Λ is 1-1 on F, Λ must be 1-1 on every F_k, hence, for every k,

$$\dim F_k = \dim \Lambda(F_k) \le \dim \Lambda X_{\le k} = \operatorname{rank} \Lambda X_{\le k}.$$

Moreover, equality is possible here for every k; it is achieved by the subspace F spanned by the monomials corresponding to bound columns in $\Lambda[X_0, X_1, \ldots]$. This proves the following

Corollary 6.2. *Assume that (Λ, F) is correct. Then, P_F is degree-reducing if and only if*

$$\dim F_k = \operatorname{rank} \Lambda X_{\leq k}, \qquad k = 0, 1, 2, \ldots. \tag{6.3}$$

In other words, if $V = [V_0, V_1, \ldots]$ is a **graded** basis for F, i.e., $V_{\leq k}$ is a basis for F_k for every k, then P_F cannot be degree-reducing unless

$$\#V_{\leq k} = \operatorname{rank} \Lambda X_{\leq k}, \qquad k = 0, 1, 2, \ldots. \tag{6.4}$$

In terms of any segmented row-echelon form $R = (R_{ij})$ for $\Lambda[X_0, X_1, \ldots]$, (6.4) is, by (ii) of Corollary 5.2, equivalent to

$$\#V_k = \#R'_{kk}, \qquad k = 0, 1, 2, \ldots.$$

Note that consideration of a graded basis for F imposes no restriction since such a basis can always be constructed inductively, starting with $k = -1$: With a basis $V_{\leq k}$ for F_k already in hand, a requisite V_{k+1} is obtained by completing $V_{\leq k}$ to a basis for F_{k+1}. For the proof of Result 6.1, we require the evident fact that we are free to include in V_{k+1} any one element from $F_{k+1} \backslash F_k$.

Proof of Result 6.1 If P_F is degree-reducing and G is any polynomial space correct for Λ, then P_F is 1-1 on G (since, by Lemma 3.2, $P_F = V(\Lambda V)^{-1}\Lambda$ with both V and $(\Lambda V)^{-1}$ 1-1, and also Λ is 1-1 on G), therefore, for any k, $\dim G_k = \dim P_F(G_k) \leq \dim F_k$, since $P_F(G_k) \subseteq P_F(\Pi_k) \subseteq \Pi_k \cap F = F_k$ (using the fact that $P_F(\Pi_k) \subseteq \Pi_k$ since P_F is degree-reducing).

Conversely, if P_F fails to be degree-reducing, then there exists g with $k := \deg g < \deg P_F g =: k'$. Since $P_F g \in F_{k'} \backslash F_{k'-1}$, we may (by the last sentence before this proof) include it in a graded basis V for F. Let U be the column map obtained from V by replacing $P_F g$ by g. Then $\Lambda V = \Lambda U$, hence also ΛU is invertible and therefore also $G := \operatorname{ran} U$ is correct for Λ. However, $\dim G_k > \dim F_k$. □

We are now ready to discuss use of an appropriately segmented row-echelon form

$$R = (R_{ij}) = M^{-1}\Lambda[X_0, X_1, \cdots]$$

for ΛX in checking whether a polynomial space F satisfying (6.3) is correct for interpolation at Λ. If V is any basis for F, then (Λ, F) is correct iff the matrix ΛV is invertible. With $V =: XW$, this is equivalent to having RW invertible (since M is invertible by assumption). Now assume that $V = [V_0, V_1, \ldots]$ is a graded basis for F. Let

$$V_k =: XW_k, \quad k = 0, 1, \ldots.$$

Then $V_j = \sum_{i\le j} X_i W_{ij}$, all i, j, i.e., the column map W, from $\mathrm{I\!R}^n$ into $\mathrm{dom}\, X$, given by $V =: XW$, is 1-1 and block upper triangular, with W_{ij} having $\#X_i$ rows and $\#V_j$ columns. Since (6.3) holds, $\#V_j = \#R'_{jj}$, all j. Hence, the matrix RW is block upper triangular, with *square* diagonal blocks. Therefore, RW is invertible if and only if $R_{kk}W_{kk}$ is invertible for each k. We have proved the following.

Proposition 6.5. *Let MR be any factorization for $\Lambda[X_0, X_1, \ldots]$ with M invertible and $R = (R_{ij})$ a correspondingly segmented row-echelon form. Assume that the polynomial space F satisfies (6.3), and let $V = XW$ be a graded basis for F, hence $W = (W_{ij})$ is block upper triangular, with W_{ij} having $\#X_i$ rows and $\#R'_{jj}$ columns. Then (Λ, F) is correct if and only if $R_{kk}W_{kk}$ is invertible for every k.*

Note that this condition only involves the diagonal blocks of W. This suggests that we choose all off-diagonal blocks of W to be zero, thus keeping the structure of F simple. This is equivalent to the requirement that

$$F = \bigoplus_k X_k W_{kk},$$

which, in turn, is equivalent to having F be scale-invariant.

Note further that the condition of the Proposition can always be met, since, by construction, each R_{kk} is onto, hence has right inverses.

If R_{kk} is square, then invertibility of $R_{kk}W_{kk}$ implies that W_{kk} is invertible, hence

$$\mathrm{ran}\, X_k W_{kk} = \mathrm{ran}\, X_k = \Pi^0_k.$$

In other words, there is then no real choice; regardless of how W_{kk} is chosen, F contains all homogeneous polynomials of degree k.

If R_{kk} is not square, then the invertibility of $R_{kk}W_{kk}$ does not completely determine the kth homogeneous component of F. How is one to choose from among the infinitely many possibilities?

The choice

$$W_{kk} = R'_{kk}, \qquad k = 0, 1, 2, \ldots \tag{6.6}$$

suggests itself, as it guarantees that $R_{kk}W_{kk}$ is invertible (since R_{kk} is onto). Moreover, if R'_{kk} was chosen as an orthonormal basis, then the diagonal block $R_{kk}W_{kk}$ becomes an identity matrix, thus facilitating the calculation (by backsubstitution) of the coefficient vector c of the resulting interpolant Xc from F to the given g. Indeed, the interpolant is of the form XWa, hence a solves the linear system

$$\Lambda XW? = \Lambda g,$$

and therefore solves the equivalent linear system

$$RW? = M^{-1}\Lambda g. \tag{6.7}$$

From a, the coefficient vector c in the power form Xc for the interpolant is obtained as

$$c = Wa.$$

The resulting interpolation scheme P_F has additional nice properties. However, we don't bother to derive them here since P_F has a major flaw: even if $\Lambda : g \mapsto g|_\Theta$, P_F depends on the coordinate system chosen in $\mathbb{R}^d$, i.e., it fails to have the desirable property that

$$P_F T g = T P_F g$$

for all $g \in \Pi$ and all 'reasonable' changes of variables, T.

7. Example

Suppose that $\Lambda : g \mapsto g|_\Theta$ with $\Theta = \{s_1\theta, \ldots, s_n\theta\}$ for some nontrivial θ and some (scalar) n-set $\{s_1, \ldots, s_n\}$. Then, for each k,

$$\Lambda X_k = [s_1^k, \ldots, s_n^k]' \, [\theta^\alpha : |\alpha| = k].$$

This shows that each ΛX_k has rank 1. Therefore, regardless of how the segmented row-echelon form $R = (R_{ij})$ for $\Lambda[X_0, X_1, \ldots]$ is obtained, each R_{kk} has just one row, and this row is a scalar multiple of the vector $(\theta^\alpha : |\alpha| = k)$. Consequently, $V_k = X_k W_{kk}$ has just one column, namely the polynomial

$$p_k := \sum_{|\alpha|=k} ()^\alpha \theta^\alpha.$$

This is unfortunate, for the following reason. If $\theta = \mathbf{i}_i$ for some i, then F consists of all polynomials of degree $< n$ in the ith indeterminate only, and that is good. However, if θ has more than one nonzero entry, then F usually differs from the natural choice in that case, namely the space of all polynomials of degree $< n$ which are constant in any direction perpendicular to θ. For F to be that space, we would need to choose each p_k as some scalar multiple of

$$\sum_{|\alpha|=k} ()^\alpha \theta^\alpha / \alpha!,$$

e.g., as the polynomial

$$x \mapsto (\theta \cdot x)^k.$$

This means that, instead of $W_{kk} = R'_{kk}$, we should choose

$$W_{kk} = \Omega_k^{-1} \, R'_{kk}, \tag{7.1}$$

with

$$\Omega_k := \operatorname{diag}(\alpha! : |\alpha| = k). \tag{7.2}$$

The space F resulting from this choice is the **least choice**, i.e., the space

$$\Pi_\Lambda = \bigoplus_k \operatorname{ran} X_k \Omega_k^{-1} R'_{kk} =: \operatorname{ran}[V_0, V_1, V_2, \ldots] \tag{7.3}$$

of de Boor and Ron [8]. In particular, it is the space Π_Θ of [5] in case $\Lambda = \ |_\Theta$, i.e., $\Lambda : g \mapsto g|_\Theta$. Correspondingly, we denote by

$$P_\Lambda, \quad \text{resp.} \quad P_\Theta$$

the resulting interpolation projectors.

To be sure, [8] arrived at this choice in a completely different way, as the linear span of all the least terms of elements of a certain space of formal power series. To make the connection between (7.3) and the definition of Π_Λ in [8], let v_i be one of the columns of $[V_0, V_1, V_2, \ldots]$. Then v_i is of the form

$$v_i = \sum_{|\alpha|=k} ()^\alpha/\alpha!\, R(i,\alpha)$$

for some k and, for that k, $R(i,\alpha) = 0$ for all $|\alpha| < k$. This implies that $v_i = f_{i\downarrow}$ (see the Introduction for the definition of $f_\downarrow$), with f_i the formal power series

$$f_i := \sum_\alpha ()^\alpha/\alpha!\, R(i,\alpha).$$

Further, $R = M^{-1}\Lambda X$, for some (invertible) M. Consequently, the f_i form a basis for the space G spanned by the formal power series

$$g_j := \sum_\alpha ()^\alpha/\alpha!\, \lambda_j ()^\alpha, \quad j = 1, \ldots, n,$$

with λ_j the jth row of the data map Λ. Further, if g is any element of G with $\deg g_\downarrow = k$, then, by Lemma 5.1(ii), $g_\downarrow$ is necessarily in the range of $X_k \Omega_k^{-1} R'_{kk}$. So, altogether, this shows that the right-hand side of (7.3) is, indeed, the linear span of all the leasts of elements of G, hence equal to Π_Λ as defined in [8].

Taking, in particular, $\Lambda = |_\Theta$, hence $\lambda_j = \delta_\theta$ (i.e., point evaluation at θ) for some $\theta \in \Theta$, we get

$$g_j = \sum_\alpha ()^\alpha/\alpha!\, \theta^\alpha = \sum_\alpha (\cdot\theta)^\alpha/\alpha! \ = \ e_\theta,$$

thus showing that, for this case, (7.3) does, indeed, give the space spanned by all the leasts of linear combinations of exponentials e_θ with $\theta \in \Theta$.

It has been pointed out (by Andreas Felgenhauer of TU Dresden, after a talk on this material) that the simple choice (6.6) actually provides (7.1) if X_k were changed to

$$X_k := [x \mapsto \prod_i x(\beta(i)) : \beta \in \{1,\ldots,d\}^k\ ],$$

i.e., if the monomials were treated as if the indeterminates did not commute, with a corresponding change of Ω_k, to the scalar matrix $k!$.

8. The least choice

The least choice, (7.3), not only provides the 'right' space in case Θ lies on a straight line, it has a rather impressive list of properties which are detailed in [5,7,8]. Of these, the following is perhaps the most unusual. Its statement uses the notation $p_\uparrow$, for the *leading* term of the polynomial p, as introduced in the Introduction. Further, if $p = Xc$, then

$$p(D) := \sum_\alpha D^\alpha c(\alpha),$$

with

$$D^\alpha = D_1^{\alpha(1)} \cdots D_d^{\alpha(d)},$$

and D_i the derivative with respect to the ith argument.

Result 8.1 ([7]).

$$\Pi_\Theta = \bigcap_{p|_\Theta = 0} \ker p_\uparrow(D).$$

A proof is provided in [8], as part of a more general argument. A direct proof can be found in [4].

It is informative to consider this result in the light of elimination, as a further motivation for the choice (7.1) for the W_{kk}. For this, we continue to denote by MR any factorization of $\Lambda[X_0, X_1, \ldots]$ into an invertible M and a correspondingly segmented row-echelon form.

Lemma 8.2. *Let $p \in \Pi_k \backslash \Pi_{<k}$ with $p_\uparrow =: X_k c$. Then, there exists $q \in \Pi_{<k}$ with $\Lambda(p-q) = 0$ if and only if $c \perp \operatorname{ran} R'_{kk}$.*

Proof. Suppose that $c \perp \operatorname{ran} R'_{kk}$. Then $R_{kk}c = 0$, hence, by (iii) of Lemma 5.1, there exists b so that $(b,c) \in \ker R_{\le k}$. This implies that $(b,c) \in \ker \Lambda X_{\le k}$, i.e., $p := X_{\le k}(b,c)$ is in $\ker \Lambda$.

Conversely, if p is of (exact) degree k and in $\ker \Lambda$, then $p = X_{\le k}(b,c)$ for some b, c with $p_\uparrow = X_k c$, and $0 = \Lambda p = \Lambda X_{\le k}(b,c) = MR_{\le k}(b,c)$, hence, since M is invertible, also $R_{\le k}(b,c) = 0$, in particular $R_{kk}c = 0$, i.e., $c \perp \operatorname{ran} R'_{kk}$. □

This lemma characterizes $\operatorname{ran} R'_{kk}$ as the orthogonal complement of the set

$$\{c : \exists\{p \in \ker \Lambda\}\ p_\uparrow = X_k c\}$$

of leading coefficients of polynomials of exact degree k in the kernel of Λ. Since

$$\Pi^0_{\Lambda,k} := \Pi_\Lambda \cap \Pi^0_k = X_k \Omega_k^{-1} R'_{kk},$$

this provides the following characterization of the kth homogeneous component of Π_Λ, i.e., of $\Pi^0_{\Lambda,k}$.

Proposition 8.3.

$$\Pi^0_{\Lambda,k} = \Pi^0_k \cap \bigcap\{\ker p_\uparrow(D) : p \in \Pi_k \backslash \Pi_{<k}, \Lambda p = 0\}.$$

Proof. Assume that q is a *homogeneous* polynomial of degree k, i.e., $q =: X_k c$, and consider the corresponding differential operator $q(D)$. Then $\Pi_{<k} \subset \ker q(D)$ trivially. Hence, for any $g \in \Pi_k \backslash \Pi_{<k}$, setting $g_\uparrow =: X_k a$, we have

$$q(D)g = q(D)(g_\uparrow) = \sum_{|\alpha|=k} c(\alpha)\alpha! a(\alpha).$$

In particular, with Ω_k as given in (7.2),

$$q(D)g = 0 \quad \Longleftrightarrow \quad c \perp \Omega_k\, a.$$

If now g is an element of Π_Λ, then $g_\uparrow = X_k\, \Omega_k^{-1} R'_{kk} b$ for some b, therefore

$$\Omega_k a = R'_{kk} b.$$

Consequently, $\Pi_\Lambda \cap \Pi_k \subset \ker q(D)$ if and only if $c \perp \operatorname{ran} R'_{kk}$, i.e., if and only if $R_{kk} c = 0$, i.e., by Lemma 8.2, if and only if q is the leading term of some polynomial in $\ker \Lambda$. □

Note that, for the case $\Lambda = \ |_\Theta$, Result 8.1 is much stronger than Proposition 8.3. For, in this special case, not only is $\Pi^0_{\Theta,k}$ annihilated by $p_\uparrow(D)$ in case $p \in \Pi_k \backslash \Pi_{<k}$ and $p|_\Theta = 0$, but *all of* Π_Θ is annihilated by such $p_\uparrow(D)$. In particular,

$$p_\uparrow(D) X_j \Omega_j^{-1} R'_{jj} = 0$$

for any j.

The reason for this much stronger result in the case $\Lambda = \ |_\Theta$ is that, in this case, $\ker \Lambda$ is a polynomial ideal and, consequently, Π_Λ is D-invariant, i.e., $f \in \Pi_\Lambda$ and $\alpha \in \mathbb{Z}_+^d$ implies $D^\alpha f \in \Pi_\Lambda$. See [8] for details.

9. Construction of an H-basis for a polynomial ideal with finite variety

Customarily, a polynomial ideal $\mathcal{I}$ is specified by some finite generating set for it, i.e., by describing it as

$$\mathcal{I} \;=\; \sum_{g\in G} \Pi g$$

for some finite $G \subset \Pi$. Given this description, it is nontrivial to determine whether or not a given $f \in \Pi$ belongs to $\mathcal{I}$, except when G is an **H-basis** for $\mathcal{I}$, meaning that, for every k,

$$\mathcal{I}_k \;:=\; \mathcal{I} \cap \Pi_k \;=\; \sum_{g\in G} \Pi_{k-\deg g} g.$$

Such a generating set is also called a **Macaulay basis** or a **canonical basis** for $\mathcal{I}$.

Indeed, if G is an H-basis for $\mathcal{I}$, then, with

$$\mathcal{I}^0 \;:=\; \{p_\uparrow : p \in \mathcal{I}\}$$

the **associated homogeneous ideal**, we have

$$\mathcal{I}^0_k := \mathcal{I}^0 \cap \Pi^0_k \;=\; \sum_{g\in G} \Pi^0_{k-\deg g} g_\uparrow, \quad \text{all } k.$$

Thus, $f \in \mathcal{I}$ if and only if

$$f_\uparrow = \sum_{g\in G} p_g g_\uparrow$$

for some $p_g \in \Pi^0_{k-\deg g}$, and in addition, for these p_g,

$$f - \sum_{g\in G} p_g g \;\in\; \mathcal{I}.$$

Since the last condition involves a polynomial of degree $< \deg f$, this leads to a terminating recursive check.

If an ideal is not given in terms of an H-basis, then it is, in general, nontrivial to construct an H-basis, except when the direct summands $\mathcal{I}^0_k$ are known in the sense that bases are known for them.

In that situation, one can construct an H-basis G for $\mathcal{I}$ by constructing

$$G_k := G \cap \Pi_k$$

inductively, i.e., for $k = 0, 1, 2, \ldots$, so that

$$\mathcal{I}_k \;=\; \sum_{g\in G_k} \Pi_{k-\deg g} g, \tag{9.1}$$

as follows.

Assume that we already have G_{k-1} in hand (as is surely the case initially for any k with $\mathcal{I}^0_{k-1} = \{0\}$, e.g., for $k = 0$). We claim that the choice

$$G_k := G_{k-1} \cup B,$$

with $B \subset \mathcal{I}_k$ so that $[B_\uparrow]$ is a basis for a linear subspace in $\mathcal{I}^0_k$ complementary to

$$\sum_{g \in G_{k-1}} \Pi^0_{k-\deg g} g_\uparrow,$$

does the job, i.e., satisfies (9.1).

Indeed, if $p \in \mathcal{I}_k$, then $p_\uparrow \in \mathcal{I}^0_k$, hence

$$p_\uparrow = [B_\uparrow]a + \sum_{g \in G_{k-1}} p_g g_\uparrow$$

for some $a \in \mathbb{C}^B$ and some $p_g \in \Pi^0_{k-\deg g}$, all $g \in G_{k-1}$. However, then

$$p - [B]a - \sum_{g \in G_{k-1}} p_g g \quad \in \quad \mathcal{I}_{k-1} = \sum_{g \in G_{k-1}} \Pi_{k-1-\deg g} g,$$

therefore

$$p \in \sum_{g \in B} \Pi_0 g + \sum_{g \in G_{k-1}} \Pi_{k-\deg g} g,$$

which was to be shown.

Note that the H-basis constructed is minimal in the sense that any proper subset of it would generate a proper subideal of $\mathcal{I}$.

An appropriate B can be constructed by: (i) starting with C for which $[C_\uparrow]$ is a basis for $\mathcal{I}^0_{k-1}$ (as constructed in the preceding step), (ii) selecting from the columns of $[()^{\mathbf{i}_j} c : c \in C, j = 1, \ldots, d]$ a column map $\tilde{C}$ which is maximal with respect to having $[\tilde{C}_\uparrow]$ 1-1, (iii) extending this to a column map $[\tilde{C}, B]$ with ($B \subset \mathcal{I}$ and) $[\tilde{C}_\uparrow, B_\uparrow]$ a basis for $\mathcal{I}^0_k$.

Of course, this requires that one have in hand, for each k, a $B_k \subset \mathcal{I}$ for which $[B_{k\uparrow}]$ is a basis for $\mathcal{I}^0_k$, and it is in this sense that one needs to "know" the $\mathcal{I}^0_k$. This information is easy to derive in case $\mathcal{I}$ is an ideal with finite codimension or, what is the same, with finite variety.

It is well-known (see, e.g., de Boor and Ron [6] for a detailed retelling of the relevant facts) that a polynomial ideal has finite codimension exactly when it is of the form $\ker \Lambda$, with $\Lambda' =: [\lambda_1, \ldots, \lambda_n]$ a basis for a subspace of Π' of the form

$$\sum_{\theta \in \Theta} \delta_\theta P_\theta(D) = \{\sum_{\theta \in \Theta} \delta_\theta p_\theta(D) : p_\theta \in P_\theta\} \tag{9.2}$$

with Θ, the variety of the ideal, a *finite* subset of $\mathbb{C}^d$, $\delta_\theta : p \mapsto p(\theta)$ the linear functional of evaluation at θ, and each P_θ a D-invariant finite-dimensional linear subspace of Π. To put it differently, a correct polynomial interpolation scheme with data map $\Lambda \in L(\Pi, \mathbb{R}^n)$ is ideal (in the sense of Birkhoff [1]) if and only if it is **Hermite interpolation**, i.e., its set of interpolation conditions is of the form (9.2).

Of course, if $\mathcal{I} = \ker \Lambda$ for some explicitly known map $\Lambda \in L(\Pi, \mathbb{R}^n)$, then membership in $\mathcal{I}$ of a given $f \in \Pi$ is trivially testable: simply compute Λf. On the other hand, an H-basis for such an ideal is likely to have its uses in the construction of error formulæ for the associated polynomial interpolation schemes.

In any case, if $\mathcal{I} = \ker \Lambda$ for some onto $\Lambda \in L(\Pi, \mathbb{R}^n)$, then

$$\mathcal{I} = X \ker R,$$

with

$$\Lambda X = MR$$

any particular factorization of ΛX with a 1-1 M.

In particular, let $R = (R_{ij})$ be a segmented row-echelon form, segmented corresponding to the segmentation $X = [X_0, X_1, \ldots]$. By Lemma 8.2, $p \in \Pi_k^0$ is in $\mathcal{I}_k^0$ iff $c := X_k^{-1} p \in \ker R_{kk}$. Consequently, any basis K_k for $\ker R_{kk}$ provides a basis, $X_k K_k$, for $\mathcal{I}_k^0$. Since $R = (R_{ij})$ is in segmented row-echelon form, each R_{jj} is onto while $R_{ij} = 0$ for $i > j$; therefore we can find $K_0, \ldots, K_{k-1}$ so that $\sum_{j<k} R_{ij} K_j = -R_{ik} K_k$ for all $i < k$, while this holds trivially for all $i \geq k$. This implies that $B_k := \sum_{j \leq k} X_j K_j$ is a column map into $\mathcal{I}$, with $B_{k\uparrow}$ a basis for $\mathcal{I}_k^0$.

If the rows of R_{kk} are, as in [7], constructed as an orthonormal basis (of the row space of a certain matrix and with respect to some inner product), then a convenient choice for K_k is the completion $[R'_{kk}, K_k]$ of R'_{kk} to an orthonormal basis for all of $\mathbb{R}^{\# R_{kk}}$.

10. The construction of rules

The **least rule for** $\mu \in \Pi'$ **from** $\Lambda \in L(\Pi, \mathbb{R}^n)$ is, by definition, the linear functional

$$\mu P_\Lambda.$$

Since $P_\Lambda = V(\Lambda V)^{-1} \Lambda$ (with V some basis for Π_Λ; see Lemma 3.2), we can also write the least rule explicitly in terms of the rows λ_i of Λ as

$$\mu P_\Lambda = a * \Lambda = \sum_i a(i) \lambda_i, \tag{10.1}$$

with

$$a := \mu V (\Lambda V)^{-1}.$$

Since our basis $V = [V_0, V_1, \ldots] =: XW$ for Π_Λ has the simple segments

$$V_k = X_k \Omega_k^{-1} R'_{kk}, \quad k = 0, 1, 2, \ldots,$$

(see (7.3)) it is not hard to apply the inverse of the matrix $(\Lambda V)' = (\Lambda XW)' = (MRW)'$ to the vector μXW. In fact, the resulting work can be carried out right along with the construction of the factorization if one is willing to record M^{-1} in some convenient form, as we now show.

The idea is to apply the program described in [7], not just to the matrix ΛX but to the 'augmented' matrix $[\Lambda; \mu]X$. (The choice of the word 'augmented' here is quite deliberate; it stresses the fact that, what is about to be described, is nothing but the dual of the standard procedure of applying Gauss elimination to the augmented matrix $[A, b]$ in order to compute a solution of $A? = b$.) For this to work, this additional, last row must never be used as pivot row. To recall, the program in [7] constructs each R'_{kk} as an orthogonal basis for the row space of a certain working array, call it B_k, whose rows consist of the X_k-part of each row not yet used as pivot row. In other words, the columns of B_k, i.e., the entries in each row of B_k, are naturally indexed by α with $|\alpha| = k$. Orthogonality is with respect to the inner product

$$\langle b, c\rangle_k := b \cdot \Omega_k^{-1} c. \tag{10.2}$$

The construction consists of making all rows in B_k not yet used as pivot rows orthogonal to those being used as pivot rows. Without that additional last row, work on B_k ceases once all rows not yet used as pivot rows are zero. With that additional row, work on this segment still ceases when all rows not used as pivot rows are zero, *except* for that additional row. This may leave that additional row nonzero, but it is certain to leave it orthogonal to the rows of the resulting R_{kk} with respect to the inner product (10.2).

At the end, we obtain the factorization

$$\begin{bmatrix} \Lambda \\ \mu \end{bmatrix} X = \begin{bmatrix} M & 0 \\ m & 1 \end{bmatrix} \begin{bmatrix} R \\ u \end{bmatrix},$$

with $\Lambda X = MR$ the earlier factorization, and with u the segmented sequence $(u_0, u_1, \ldots)$ such that

$$u_k \cdot \Omega_k^{-1} R'_{kk} = 0$$

for every nonempty R_{kk}. One verifies that

$$\begin{bmatrix} M & 0 \\ m & 1 \end{bmatrix}^{-1} = \begin{bmatrix} M^{-1} & 0 \\ -mM^{-1} & 1 \end{bmatrix}.$$

(Here, m is a $1 \times (m-1)$-matrix rather than a sequence, hence I write mM^{-1} rather than the (undefined) $m \cdot M^{-1}$.) This implies that

$$(-mM^{-1}\Lambda + \mu)V = (-mM^{-1}\Lambda + \mu)XW = \sum_k u_k \cdot \Omega_k^{-1} R'_{kk} = 0.$$

Consequently, $mM^{-1}\Lambda V = \mu V$, hence the desired coefficient vector a for (10.1) equals the one row of mM^{-1}.

11. A second look at Gauss elimination by segments

Here is a second look at Gauss elimination by segments, from the point of view of the data map $\Lambda \in L(\Pi, \mathbb{R}^n)$. This discussion is applicable to any (segmented) matrix $A = [A_0, A_1, \ldots]$ since an arbitrary matrix $A \in \mathbb{F}^{n \times J}$ can be thought of as such a product ΛX of a row map $\Lambda \in L(S, \mathbb{F}^n)$ with a column map $X \in L(\mathbb{F}^J, S)$ in at least two ways:

(1) $S = \mathbb{F}^n$ and $\Lambda = \mathrm{id}$, hence $X = A = [A(:,j) : j \in J]$.

(2) $S = \mathbb{F}^J$ and $V = \mathrm{id}$, hence $\Lambda = A \in L(S, \mathbb{F}^n)$.

At the same time, it is to be hoped that elimination by segments will be found of help in other applications where it is more natural to do elimination by certain segments rather than by columns. For example, a process requiring a certain amount of column pivoting might be a good prospect.

For this reason, in this section, $X = [X_0, X_1, \ldots]$ is any segmented column map, from $\mathbb{F}_0^J := (J \to \mathbb{F})_0$ into some linear space S, with J some set, and Λ is any onto data map or **row map** on S, and the matrix of interest is

$$A := \Lambda X = [\Lambda X_0, \Lambda X_1, \ldots] =: [A_0, A_1, \ldots],$$

as before.

We have already made use of the following fact: with $\lambda_1, \ldots, \lambda_n$ the rows of the data map $\Lambda \in L(S, \mathbb{F}^n)$, the dual map $\Lambda' \in L(\mathbb{F}^n, S')$ for Λ is of the form

$$\Lambda' = [\lambda_1, \ldots, \lambda_n] : \mathbb{F}^n \to S' : a \mapsto \sum_i a(i)\lambda_i.$$

Since Λ is onto by assumption, Λ' is 1-1, hence a basis for its range.

For each k, let Λ_k be a data map whose rows, when restricted to $S_k := \operatorname{ran} X_{\le k}$, provide a basis for the linear space

$$L_k := \{\lambda|_{S_k} : \lambda \in L,\ \lambda X_{<k} = 0\}.$$

Equivalently, Λ_k is any onto data map with its $\dim L_k$ rows taken from $\operatorname{ran}\Lambda'$, for which $\Lambda_k X_{<k} = 0$ while $\Lambda_k X_k$ is onto.

However this is actually done numerically, we will have constructed a basis

$$[\Lambda_k' : k = 1, 2, \ldots]$$

for $\operatorname{ran}\Lambda'$. As Λ' is 1-1, hence a basis for its range, this means that we have, in effect, constructed an invertible matrix $M' \in \mathbb{R}^{m \times m}$ so that

$$\Lambda' = [\Lambda_k' : k = 0, 1, 2, \ldots]M'.$$

Thus,

$$\Lambda X = A = MR,$$

with

$$R := [\Lambda_0; \Lambda_1; \Lambda_2; \ldots] [X_0, X_1, X_2, \ldots] =: (R_{ij} : i, j = 0, 1, 2, \ldots)$$

a block upper triangular matrix since

$$R_{ij} = \Lambda_i X_j$$

is trivial for $i > j$.

Consider now, in particular, the diagonal blocks,

$$R_{kk} = \Lambda_k X_k.$$

If λ is any linear functional in $\operatorname{ran}\Lambda'$ which vanishes on $\Pi_{<k}$, then there is exactly one coefficient sequence c so that $\lambda = \Lambda_k' c$ on Π_k. In other words, the rows of R_{kk} provide a basis for the linear subspace $\{(\lambda()^\alpha : |\alpha| = k) : \lambda \in \operatorname{ran}\Lambda', \lambda X_{<k} = 0\}$. In particular, R_{kk}' is 1-1, and its range is independent of the particular choice of the map Λ_k'. This recovers (i) of Corollary 5.2 in this more general setting.

Acknowledgements I acknowledge with pleasure and gratitude various extended discussions I had with Nira Dyn concerning the material of this paper, particularly during a visit to Israel, in February 1993, funded by the US-Israeli Binational Science Foundation.

The work was supported by the National Science Foundation under Grant No. DMS-9224748, by the United States Army under Contract DAAL03-90-G-0090, by the US-Israel Binational Science Foundation under Grant No. 90-00220, and by the Alexander von Humboldt Stiftung.

References

[1] Birkhoff G. The algebra of multivariate interpolation. In *Constructive approaches to mathematical models*, pages 345–363. Academic Press (New York), 1979. C.V. Coffman and G.J. Fix, eds.

[2] de Boor C. Polynomial interpolation in several variables. In *Proceedings of the Conference honoring Samuel D. Conte*, pages xxx–xxx. Plenum Press (New York), 199x. R. DeMillo and J.R. Rice, eds.

[3] de Boor C. An alternative approach to (the teaching of) rank, basis and dimension. *Linear Algebra Appl.*, 146:221–229, 1991.

[4] de Boor C. On the error in multivariate polynomial interpolation. *Applied Numerical Mathematics*, 10:297–305, 1992.

[5] de Boor C., Ron A. On multivariate polynomial interpolation. *Constr. Approx.*, 6:287–302, 1990.

[6] de Boor C., Ron A. On polynomial ideals of finite codimension with applications to box spline theory. *J. Math. Anal. Appl.*, 158:168–193, 1991.

[7] de Boor C., Ron A. Computational aspects of polynomial interpolation in several variables. *Math. Comp.*, 58:705–727, 1992.

[8] de Boor C., Ron A. The least solution for the polynomial interpolation problem. *Math. Z.*, 210:347–378, 1992.

[9] Lorentz R. A. *Multivariate Birkhoff Interpolation*. volume 1516 of *Lecture Notes in Mathematics*, Springer-Verlag (Heidelberg), 1992.