What is the inverse of a basis?

Carl de Boor

Abstract.

A formula is given for the inverse of the linear map, from coordinate space to a linear space, induced by a basis for that linear space, which is then connected to various Applied Linear Algebra constructs.

AMS subject classification: 65F20.

Key words: Applied Linear Algebra, interpolation, polynomials, splines.

1. Introduction

A paper on something as basic as the basis of a vector space is not likely to offer some new mathematics, and this paper is no exception. Rather, it is meant to illustrate that a certain point of view regarding bases leads to clarity and efficiency in their use.

In its extreme form (see, e.g., [B]), this point of view is the following. Since the sole purpose of a basis of a vector space over the field $\mathbb{F}$ ($:= \mathbb{R}$ or $\mathbb{C}$) is to provide an invertible linear map to that vector space from some coordinate space $\mathbb{F}^n$, it seems most direct to *define* bases that way, i.e., as invertible linear maps to the vector space from some $\mathbb{F}^n$ or, more generally, from $\mathbb{F}^I$ for some index set I.

However, it is very hard to effect a change in as time-honored a concept as that of a basis. My more modest ambition therefore is to help shift attention, from the basis to the invertible linear map specified by it. Still, it is good to have a name for such a map. Such a map is called a *frame* in Differential Geometry, but 'frame' is also used in Harmonic Analysis with a related but different meaning. So, wary of calling such a map a basis, I will call it here, more modestly, a **basis-map**.

Any linear map V from $\mathbb{F}^n$ to the vector space X over $\mathbb{F}$ is necessarily of the form

$$V : \mathbb{F}^n \to X : a \mapsto \sum_{j=1}^{n} \mathbf{v}_j a_j$$

for a certain sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in X that comprises the images under V of the unit vectors

$$e_j := (\underbrace{0, \ldots, 0}_{j-1}, 1, 0, \ldots), \quad j = 1, \ldots, n.$$

In view of the special case $X \subset \mathbb{F}^m$, it is therefore natural to write such V explicitly as

$$V = [\mathbf{v}_1, \ldots, \mathbf{v}_n],$$

and call this a **column map** into X, and call $\mathbf{v}_j$ its jth **column**. This terminology is further supported by the fact that, for any linear map A whose domain includes X,

$$AV = [A\mathbf{v}_1, \ldots, A\mathbf{v}_n].$$

In these terms, a basis-map is an invertible column map into X; the sequence of its columns is the corresponding basis. More generally, for any column map, the linear combinations of its column sequence make up the range of the map; the sequence is linearly independent exactly when the map is 1-1; the sequence is spanning for X exactly when the map is onto.

With this, it is natural to ask for the *inverse* of a basis-map $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ for X. If X is itself a coordinate space, hence necessarily $X = \mathbb{F}^n$, then V is (or has a natural representation as) a matrix, and V^{-1} is simply that, the matrix inverse to V. In any other case, it is, offhand, not obvious how to obtain V^{-1}. It is the purpose of this note to point out a simple recipe which, not surprisingly, is implicit in all constructions of V^{-1} in the literature and therefore, by rights, should be explicitly in every (Applied) Linear Algebra textbook, to discuss briefly the connection of this formula to certain basic constructs in Applied Linear Algebra, and to give some other examples for the efficacy of this point of view.

2. Motivation: Change of basis and matrix representation

Since most textbooks on Linear Algebra define a basis as a set (not even a sequence, but a set!) with certain properties, a change of basis is one of the more harrowing maneuvers for the students, involving, as it does, some transition matrix somehow connected to the two basis sets and leaving the students with the feeling of not knowing whether they are coming or going.

In contrast, such change of basis is immediate in terms of the corresponding basis-maps. Assume we are to switch, from the basis-map W to the basis-map V, i.e., we know the coordinates, b, of $\mathbf{x} \in X$ with respect to W, i.e., we know that $\mathbf{x} = Wb$. Then, the coordinates a of $\mathbf{x}$ with respect to V can be computed as $a = V^{-1}\mathbf{x} = V^{-1}Wb$. In particular, $V^{-1}W$ is the *transition matrix*, for going from the basis W to the basis V. It is also the matrix representation, with respect to W and V, for the identity on X.

If, more generally, T is a linear map from some vector space Z with basis-map W to the vector space X with basis-map V, then the matrix $A := V^{-1}TW$ is easily recognized as the matrix representation for T with respect to W and V, in the standard sense that the coordinate vector of $T\mathbf{z} = TWb$ with respect to the basis V is the vector Ab.

At this point into the story, some student is certain to ask for a *formula* for V^{-1}, and so provides the motivation for the present note.

3. A formula for V^{-1}

While Linear Algebra books seem silent on the question of the inverse of a basis-map $V : \mathbb{F}^n \to X$, Numerical Analysis has a straightforward answer: "discretize!", i.e., replace the abstract equation

$$V? = \mathbf{x},$$

for the V-coordinates $a := V^{-1}\mathbf{x}$ of the vector $\mathbf{x} \in X$, by the numerical equation

$$\Lambda' V? = \Lambda' \mathbf{x}, \tag{3.1}$$

with Λ' some suitable linear map from X to $\mathbb{F}^n$.

To be sure, any linear map from X to $\mathbb{F}^n$ is necessarily of the form

$$X \to \mathbb{F}^n : \mathbf{x} \mapsto (\lambda_1 \mathbf{x}, \ldots, \lambda_n \mathbf{x})$$

for some linear functionals $\lambda_1, \ldots, \lambda_n$ on X, and this sequence completely specifies the map. In light of the special case $X = \mathbb{F}^m$, it is natural to write this map

$$\Lambda' = [\lambda_1, \ldots, \lambda_n]',$$

and call it a **row map**, with **rows** $\lambda_1, \ldots, \lambda_n$.

How to get Λ'? In the typical situation, and the one I will now assume here, X is a linear subspace of some linear space Y with a ready supply of explicitly known linear functionals. For example, Y might be an inner product space, or Y might be a space of functions we can evaluate, differentiate, average, etc., hence it is easy to construct various row maps $\Lambda' : Y \to \mathbb{F}^n$.

Which Λ' is suitable? Any Λ' that makes the **Gram matrix**

$$\Lambda' V = [\lambda_1, \ldots, \lambda_n]'[\mathbf{v}_1, \ldots, \mathbf{v}_n] = (\lambda_i \mathbf{v}_j : i, j = 1{:}n)$$

invertible. In that case, directly from (3.1),

$$V^{-1}\mathbf{x} = (\Lambda' V)^{-1} \Lambda' \mathbf{x},$$

hence

$$V^{-1} = (\Lambda' V)^{-1} \Lambda'|_X \tag{3.2}$$

is the promised formula.

It follows that

$$V(\Lambda' V)^{-1} \Lambda'$$

is (a beautiful way to write) the identity on X. In other words,

$$\mathbf{x} = V(\Lambda' V)^{-1} \Lambda' \mathbf{x}$$

is the irredundant representation of $\mathbf{x} \in X$ provided by the basis-map V.

4. A simple example

The simplest situation in which (3.2) is useful occurs when X is a proper linear subspace of some coordinate space, i.e., $Y = \mathbb{F}^m$ (with $m > n$). In this case, V is commonly represented by a matrix, namely the matrix whose columns comprise the basis elements (some refer to this as the basis *matrix*, to distinguish it from the basis). Correspondingly, we think of the linear map $\Lambda' : Y = \mathbb{F}^m \to \mathbb{F}^n$ as a matrix, too, i.e., $\Lambda' \in \mathbb{F}^{n \times m}$.

A natural choice for Λ' in this case is V^*, the (conjugate) transpose of V, since V is 1-1 hence so is V^*V, hence V^*V is invertible (since it is square). The right-hand side in the resulting formula

$$V^{-1} = (V^*V)^{-1} V^* \tag{4.1}$$

is often referred to as the *generalized* inverse of V since, strictly speaking, (4.1) only holds on X.

More generally, V^{-1} agrees on X with $(\Lambda' V)^{-1}\Lambda'$ for any $\Lambda' \in \mathbb{F}^{n\times m}$ for which $\Lambda' V$ is invertible. In particular, for each such Λ',

$$P := V(\Lambda' V)^{-1}\Lambda' \tag{4.2}$$

is the identity on its range, X, hence is a *linear projector*, from $\mathbb{F}^m$ onto X. It is the unique such linear projector P for which

$$\Lambda' P\mathbf{y} = \Lambda'\mathbf{y}, \qquad \mathbf{y} \in Y = \mathbb{F}^m,$$

and any linear projector onto X arises in this way. In these terms, the particular choice $\Lambda' = V^*$ gives the *orthogonal* projector onto X.

5. Gramians, projectors, interpolation, and change of basis

In full generality, X lies in some linear space Y and, correspondingly, Λ' is some linear map from Y to $\mathbb{F}^n$ for which the Gramian matrix, $\Lambda' V$, is invertible, in which case (4.2) gives the unique linear projector, from Y onto X, for which

$$\Lambda' P\mathbf{y} = \Lambda'\mathbf{y}, \qquad \mathbf{y} \in Y.$$

If Y happens to be a space of functions, and Λ' associates $\mathbf{y} \in Y$ with its restriction $\mathbf{y}|_{\mathrm{T}} := (\mathbf{y}(\tau) : \tau \in \mathrm{T})$ to some n-set T, then, assuming that $\Lambda' V$ is invertible, we recognize in $P\mathbf{y}$ the unique interpolant to $\mathbf{y} \in Y$ from X at the points in T. The same language seems therefore appropriate in the general case, except that 'matching at the points of T' becomes, more generally, a 'matching of the information about $\mathbf{y}$ contained in $\Lambda'\mathbf{y}$'.

In this way, one recognizes the available formulas for the inverse of a basis-map to be in 1-1 correspondence with the linear projectors onto the range of the basis-map.

Yet again, we can think of this process as nothing more than a change of basis. For, if $\Lambda' V$ is invertible, then so is $\Lambda'|_X$, and its inverse, $W := (\Lambda'|_X)^{-1}$, is necessarily an invertible linear map from $\mathbb{F}^n$ to X, hence a basis-map for X. Knowing $\Lambda'\mathbf{x}$ for some $\mathbf{x} \in X$ means knowing the coordinates of $\mathbf{x}$ with respect to that basis-map W, from which our formula, $P\mathbf{x} = V(\Lambda' V)^{-1}\Lambda'\mathbf{x}$, merely constructs the coordinates $(\Lambda' V)^{-1}\Lambda'\mathbf{x}$ for $\mathbf{x}$ with respect to V, showing $(\Lambda' V)^{-1}$ to be the corresponding transition matrix.

6. Examples.

6.1 Polynomial interpolation As a standard example, take for V the **power basis**(-map) for the space $X = \Pi_{<n}$ of *polynomials of degree* $< n$, i.e., $\mathbf{v}_j(t) = t^{j-1}$, $j = 1, \ldots, n$, as a subspace of the space $Y := \mathbb{R}^{\mathbb{R}}$ of all real-valued functions on the real line. Take for Λ' the map $\Lambda' : Y \to \mathbb{R}^n : \mathbf{y} \mapsto \mathbf{y}|_{\mathrm{T}}$ that associates $\mathbf{y} \in Y$ with the n-vector of its values at the n-set $\mathrm{T} = \{\tau_1, \ldots, \tau_n\}$. Since the functions $\boldsymbol{\ell}_j(t) := \prod_{k\neq j}(t - \tau_k)/(\tau_j - \tau_k)$, $j = 1, \ldots, n$, are in $\Pi_{<n}$ and $\Lambda'[\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n] = \mathrm{id}$ by inspection, $[\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n]$ must be 1-1, hence onto (since $\Pi_{<n} = \mathrm{ran}[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is at most n-dimensional), therefore a basis-map for $\Pi_{<n}$, and Λ' must therefore be 1-1 on $\Pi_{<n}$. Consequently, also V is a basis-map for

$\Pi_{<n}$ and therefore $\Lambda' V$ is invertible, and $P\mathbf{y} = V(\Lambda' V)^{-1}\Lambda'\mathbf{y}$ is the unique polynomial of degree $< n$ that matches $\mathbf{y}$ on T, written here in **power form**, i.e., in terms of the power basis.

6.2 V^{-1} by elimination To be sure, elementary Linear Algebra does provide a recipe for the computation of the coordinates of $\mathbf{x} \in X$ with respect to any particular basis V for X in case X is some subspace of $\mathbb{F}^n$, but this recipe is not much stressed. To recall: Elimination applied to some matrix A produces its reduced row echelon form and, from this, by dropping all zero rows if any, its really reduced echelon form, $R := \mathrm{rrref}(A)$. In the process, each of the columns of A is classified as 'free' or 'bound', depending on whether or not it is in the span of the columns to the left of it. If we let `free` and `bound` denote the corresponding increasing sequences of indices, then $A(:,\mathtt{bound})$ is a basis-map for $\operatorname{ran} A$, while $A = A(:,\mathtt{bound})R$, hence, for any j, $R(:,j)$ contains the coordinates of $A(:,j)$ with respect to $A(:,\mathtt{bound})$. In particular, $R(:,\mathtt{bound}) = \mathrm{id}$.

This implies that, for any $\mathbf{x} \in X$, the last column of $\mathrm{rrref}([V,\mathbf{x}])$ contains the coordinates of $\mathbf{x}$ with respect to any basis V for X. Further, with $r := \#V$, the $(n, r{+}n)$-matrix $R := \mathrm{rrref}([V, \mathrm{id}_n])$ provides in $U := R(:, r + (1{:}n))$ the invertible matrix for which $U[V, \mathrm{id}_n] = R$. With V 1-1, the sequence `bound` in this case is sure to start with $(1{:}r)$, hence $\mathrm{id}_r = R(1{:}r, 1{:}r) = U(1{:}r, :)V$, showing $W := U(1{:}r, :)$ to be a left inverse for V, hence $\mathbf{x} = VW\mathbf{x}$ for any $\mathbf{x} \in X = \operatorname{ran} V$. Further, since $R(:,\mathtt{bound}) = \mathrm{id}_n$, it follows with $\mathtt{bound} =: [1{:}r, r + \mathtt{b}]$ that $W(:,\mathtt{b}) = 0$. Thus, with $W(:,\mathtt{f})$ the other columns of W, we have $W\mathbf{x} = W(:,\mathtt{f})\mathbf{x}(\mathtt{f})$, and, in particular, $\mathrm{id}_r = WV = W(:,\mathtt{f})V(\mathtt{f},:)$, hence $W(:,\mathtt{f}) = V(\mathtt{f},:)^{-1}$. Therefore, finally, we obtain the following special case of (3.2):

$$V^{-1}\mathbf{x} = W\mathbf{x} = W(:,\mathtt{f})\mathbf{x}(\mathtt{f}) = (\Lambda' V)^{-1}\Lambda'\mathbf{x},$$

with

$$\Lambda' : \mathbf{y} \mapsto \mathbf{y}(\mathtt{f}).$$

Note that the corresponding linear projector, $P := VW$, is interpolation in the narrow meaning of that word: For each $\mathbf{y} \in \mathbb{F}^n$, $P\mathbf{y}$ is the unique vector in $X = \operatorname{ran} V$ for which $(P\mathbf{y})(j) = \mathbf{y}(j)$ for all $j \in \mathtt{f}$.

A particularly striking consequence of these observations is the following. Suppose we are trying to find out whether the linear map A carries the linear subspace Z in its domain into the linear subspace X of its target. With W and V basis-maps for Z and X, respectively, this will happen iff $AW \subset \operatorname{ran} V$, hence, if and only if $R := \mathrm{rrref}([V, AW])$ has exactly $r := \#V$ rows, in which case, as a kind of unexpected extra, $R(:, r{+}(1{:}\#W)) = V^{-1}AW$, i.e., one obtains the matrix representation of $A|_Z : Z \to X$ with respect to the basis-maps W and V for its domain and target.

For the very special case $A = \mathrm{id}_X$, one obtains in this way the transition matrix $V^{-1}W$ for two basis-maps of a linear subspace X of $\mathbb{F}^n$.

6.3 The Jordan canonical form This example brings the essential part of Ptak's [P] concise argument for the existence of the Jordan canonical form.

Assume that the linear map A on Y to Y is nilpotent of degree q, and that $A^{q-1}\mathbf{z} \neq 0$. Then there is some linear functional μ on Y for which $\mu A^{q-1}\mathbf{z} \neq 0$, hence, with $\Lambda' := [\mu A^{q-1}, \ldots, \mu A, \mu]'$ and $V := [\mathbf{z}, A\mathbf{z}, \ldots, A^{q-1}\mathbf{z}]$, the matrix $\Lambda' V$ is triangular with nonzero diagonal entries, hence invertible. Consequently, V is 1-1, hence is a basis-map for its range, $X := \operatorname{ran} V$, and X is A-invariant, and, since $AV = [A\mathbf{z}, \ldots, A^{q-1}\mathbf{z}, \mathbf{0}]$, the matrix representation with respect to V for A restricted to X is $[e_2, \ldots, e_q, 0]$, i.e., the Jordan block of order q for the eigenvalue 0. Moreover, with $P = V(\Lambda' V)^{-1}\Lambda'$ the corresponding linear projector, $Z := \ker P = \ker \Lambda'$ is also A-invariant, hence provides the A-invariant direct sum decomposition $Y = X \oplus Z$ for Y. Therefore, if Y is minimally A-invariant, in the sense that there is no A-invariant direct sum decomposition of Y with two nontrivial summands, then V provides a basis for Y with respect to which the matrix representation for A is that Jordan block.

6.4 The basis of the trivial subspace Extreme situations nicely test one's understanding of subject matter. For bases and dimensions, this extreme situation occurs when the space in question is trivial. Almost all textbooks maintain that the trivial space has *no* basis, and then proceed to *define* its dimension to be zero. In fact, if by basis of a space we mean an invertible column map to that space, then the trivial space has exactly one basis (the only linear space with that property), namely the one and only map from $\mathbb{F}^0$ to that space, the one that maps the sole element of $\mathbb{F}^0$, namely the empty sequence, (), to the sole element of the trivial space, namely 0. This is a column map with *no* columns, reasonably denoted by [ ], and, correspondingly, the dimension of $\{0\}$ is 0. Note that [ ] is the only linear map whose condition number is 0, thus 'proving' the rule that the condition number of any linear map is ≥ 1.

6.5 Discrete least-squares cubic spline approximation in `MATLAB` The view of a basis as an invertible linear map has also quite practical benefits, as the following example is intended to illustrate.

Let X be the linear space of all C^2 cubic splines with break sequence $\xi = (\xi_1 < \cdots < \xi_{\ell+1})$, and consider the construction of $f \in X$ that minimizes

$$E(f) := \sum_j (y_j - f(t_j))^2$$

for given sequences y and t, the latter assumed to be strictly increasing, for simplicity. Given any basis-map V for X, and with

$$\Lambda_t' : f \mapsto f|_t = (f(t_j) : j),$$

the minimizing spline f is well-known to be

$$f = V(G'G)^{-1}G'y,$$

provided the Gram matrix

$$G := \Lambda_t' V$$

is 1-1. Assuming this to be the case, we could therefore construct f in `MATLAB` as $V(G\backslash y)$, provided we had some basis-map V for the spline space X in hand.

Now, the only spline construction tool provided by plain `MATLAB` is the command

$$\texttt{spline(xi,z)} \tag{6.1}$$

which returns the unique cubic spline f with break sequence $\mathtt{xi}$ that matches these data in the sense that $f(\mathtt{xi}) = \mathtt{z}(2{:}\mathtt{end}-1)$ and $Df(\min \mathtt{xi}) = \mathtt{z}(1)$, $Df(\max \mathtt{xi}) = \mathtt{z}(\mathtt{end})$. But that is almost all we need since, with $\mathtt{xi} = \xi$, hence $\mathtt{z}$ of length $n := \ell + 3$, this identifies $\mathtt{z} \mapsto \mathtt{spline}(\mathtt{xi}, \mathtt{z})$ as a 1-1 linear map from $\mathbb{R}^n$ onto X, i.e., a basis-map. Call this map V.

The spline command has two additional features. If $\mathtt{z}$ is an array of size [d,n], then (6.1) is the unique d-vector-valued C^2 cubic spline f with break sequence $\mathtt{xi}$ that satisfies

$$f(\mathtt{xi}) = \mathtt{z}(:, 2{:}\mathtt{n}-1) \text{ and } Df(\min \mathtt{xi}) = \mathtt{z}(:, 1), Df(\max \mathtt{xi}) = \mathtt{z}(:, \mathtt{n}).$$

In particular, the command spline(xi,eye(n)) returns the vector-valued spline whose jth component is the jth column of the basis-map V. Also, the command

$$\mathtt{spline}(\mathtt{xi}, \mathtt{z}, \mathtt{t})$$

returns, not the interpolating spline itself, but its values at the entries of $\mathtt{t}$. Hence, with $\mathtt{t}$ the row-vector containing the entries of t, the command

$$\mathtt{spline}(\mathtt{xi}, \mathtt{eye}(\mathtt{n}), \mathtt{t})$$

returns the transpose of the matrix $G = \Lambda'_t V$. Therefore, with $\mathtt{y}$ the column-vector corresponding to the data y, the one-line command

$$\mathtt{spline}(\mathtt{xi}, \mathtt{spline}(\mathtt{xi}, \mathtt{eye}(\mathtt{n}), \mathtt{t})' \backslash \mathtt{y})$$

provides the sought-for least-squares approximation. In fact, if $\mathtt{y}$ is the row-vector corresponding to y, then the following transposition-free statement will already do the job:

$$\mathtt{spline}(\mathtt{xi}, \mathtt{y}/\mathtt{spline}(\mathtt{xi}, \mathtt{eye}(\mathtt{n}), \mathtt{t}))$$

7. Analysis and synthesis; row maps and column maps

The maps Λ' and V play dual roles in the above discussion. With X some n-dimensional linear subspace of some vector space Y over the scalar field $\mathbb{F}$, we recognize in $\Lambda' : Y \to \mathbb{F}^n$ a **data map** or **analysis** operator, i.e., an operator that extracts numerical information from the elements of Y. Further, $V : \mathbb{F}^n \to X$ is a **synthesis** operator, i.e., an operator that constructs an element of X from numerical information.

In this way, one also recognizes

$$V\Lambda' = [\mathbf{v}_1, \ldots, \mathbf{v}_n][\lambda_1, \ldots, \lambda_n]' : \mathbf{y} \mapsto \sum_j (\lambda_j \mathbf{y})\mathbf{v}_j$$

as the general linear map on Y with range in $X = \operatorname{ran} V$.

The situation is particularly nice in case V and Λ' are **dual to each other**, i.e., $\Lambda' V = \mathrm{id}$, in which case (cf. (4.2)) $V\Lambda'$ is a linear projector onto the range of V. There is no dearth of important examples, since all standard expansions (Fourier, Taylor, Newton, wavelets, etc, etc) follow this pattern of analysis and synthesis (albeit some of these involve infinite sums).

As a standard example, the truncated Taylor expansion at 0,

$$\mathbf{y}(t) \sim \sum_{j<n} t^j \, D^j \mathbf{y}(0)/j!,$$

takes advantage of the fact that the data map $\Lambda' : \mathbf{y} \mapsto (D^j \mathbf{y}(0)/j! : j = 0, \ldots, n-1)$ is dual to the power basis-map for $\Pi_{<n}$. It provides the unique polynomial of degree $< n$ that matches the value at 0 of $\mathbf{y}$ and its first $n-1$ derivatives.

If the Gramian $\Lambda' V$ is merely invertible, then any factorization $\Lambda' V = LU$ of the Gramian into square, hence invertible, matrices L and U leads to the modified maps

$$\widehat{V} := VU^{-1}, \qquad \widehat{\Lambda}' := L^{-1}\Lambda'$$

which describe the same projector,

$$\widehat{V}\widehat{\Lambda}' = P = V(\Lambda' V)^{-1}\Lambda',$$

but more simply since they also are dual to each other.

7.1 Gram-Schmidt A basic example is, of course, Gram-Schmidt orthogonalization which, applied to the basis $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, proceeds in exactly this way, with $\lambda_i : \mathbf{y} \to \langle \mathbf{y}, \mathbf{v}_i \rangle$, all i, using Gauss elimination without pivoting to obtain an orthogonal basis $\widehat{V} = VU^{-1}$ via factorization of the Gram matrix $\Lambda' V = LU$ into a lower and an upper triangular factor, though its standard implementation does make use of the considerable latitude available in Gauss elimination in the order in which one carries out the computations, and purists would insist on the normalization $L = U^*$.

7.2 Newton form and Gauss elimination Take again for V the power basis-map for polynomials of degree $< n$, and for Λ' the map of restriction to some n-set $\mathrm{T} = \{\tau_1, \ldots, \tau_n\}$ of points, and use Gauss elimination without pivoting to factor the Gramian (named in this case, by Lebesgue, the **Vandermonde** matrix) into a lower triangular matrix L and a unit upper triangular matrix U, and thereby obtain, in $\widehat{V}\widehat{\Lambda}'$, the *Newton form* of the interpolating polynomial.

Indeed, since U is unit upper triangular, $\widehat{\mathbf{v}}_j$, as the jth column of VU^{-1}, has the leading term t^{j-1}. Further, $\Lambda'\widehat{V} = L\widehat{\Lambda}'\widehat{V} = L$ is lower triangular, hence $\widehat{\mathbf{v}}_j$ must vanish at $\tau_1, \ldots, \tau_{j-1}$. So, altogether, $\widehat{\mathbf{v}}_j = \prod_{k<j}(\cdot - \tau_k)$, $j = 1, \ldots, n$, and these are the polynomials appearing in the Newton form for the interpolating polynomial. At the same time, since L is lower triangular, $\widehat{\lambda}_j$, as the jth row of $L^{-1}\Lambda'$, is a linear combination of evaluations at $\tau_1, \ldots, \tau_j$, and, since $\widehat{\Lambda}' V = \widehat{\Lambda}'\widehat{V}U = U$ is unit upper triangular, $\widehat{\lambda}_j$ vanishes on the monomials $t^0, \ldots, t^{j-2}$ and is 1 on the monomial t^{j-1}. In other words,

$$\widehat{\lambda}_j = \Delta(\tau_1, \ldots, \tau_j),$$

the divided difference at the points $\tau_1, \ldots, \tau_j$ (as is to be expected from the jth coefficient in the Newton form, and using the very apt notation for it introduced by W. Kahan).

7.3 Divided differences Here is a final example for the efficacy of basis-maps. It concerns the smooth dependence of the divided difference $\Delta(\tau)$ on the point $\tau \in \mathbb{R}^n$, as a linear functional on the linear space Π of (univariate) polynomials, say. Here is the quickest way to this important result.

Consider the linear map

$$V_\tau : \mathbb{F}_0^{\mathbb{N}} \to \Pi : a \mapsto \sum_j a_j \mathbf{v}_{j,\tau},$$

with

$$\mathbf{v}_{j,\tau} : t \mapsto (t - \tau_1) \cdots (t - \tau_{j-1}), \quad j = 1, 2, \ldots,$$

and $\mathbb{F}_0^{\mathbb{N}}$ the collection of all infinite sequences with only finitely many nonzero entries.

The map V_τ is well-defined, linear, 1-1, and onto, hence a basis-map for Π, the so-called Newton basis. It follows that

$$V_\tau^{-1} : \Pi \to \mathbb{F}_0^{\mathbb{N}} : p \mapsto [\lambda_j : j \in \mathbb{N}]' p =: (\Delta(\tau_1, \ldots, \tau_j)p : j \in \mathbb{N})$$

exists, and, since V_τ is a smooth function of τ, so is V_τ^{-1}.

Of course, for this argument to be of any value, I need to connect this *definition* of the divided difference $\Delta(\tau_1, \ldots, \tau_j)$ with the more customary algorithmic definition. But that is easy.

First, for any $p \in \Pi$,

$$p = V_\tau V_\tau^{-1} p = \sum_{i=1}^{j} \mathbf{v}_{i,\tau} \lambda_i p \quad + \quad g \cdot \mathbf{v}_{j+1,\tau},$$

with the first term in $\Pi_{<j}$ and the factor g in the second term some polynomial, hence $\lambda_j p$ is the leading coefficient of the unique polynomial of degree $< j$ that agrees with p at $\tau_1, \ldots, \tau_j$ in the sense that their difference is divisible by $\mathbf{v}_{j+1,\tau}$, therefore λ_j is indeed only a function of $\tau_1, \ldots, \tau_j$, thus justifying the notation $\Delta(\tau_1, \ldots, \tau_j)$ for it, and is symmetric in the τ_i.

Second, evaluation by Horner's scheme shows that, for any $z \in \mathbb{R}$,

$$\sum_{j \le k} a_j \prod_{i<j} (\cdot - \tau_i) = \sum_{j \le k} \widetilde{a}_j \prod_{i<j} (\cdot - \widetilde{\tau}_i)$$

with

$$\widetilde{a}_j = a_j + \widetilde{a}_{j+1}(z - \tau_j), \quad (\widetilde{\tau}_1, \ldots, \widetilde{\tau}_k) := (z, \tau_1, \ldots, \tau_{k-1}),$$

hence

$$\Delta(z, \tau_1, \ldots, \tau_{j-1}) = \Delta(\tau_1, \ldots, \tau_j) + \Delta(z, \tau_1, \ldots, \tau_j)(z - \tau_j),$$

which is the identity underlying the algorithmic definition of the divided difference.

8. Notation

The notations $[\mathbf{v}_1,\ldots,\mathbf{v}_n]$ and $[\lambda_1,\ldots,\lambda_m]'$ for column maps and row maps, respectively, were chosen in strict adherence to `MATLAB` notation. (I abandoned an earlier way of writing the general row map as $[\lambda_1;\ldots;\lambda_n]$, using `MATLAB`'s notation for a matrix with rows $\lambda_1,\ldots,\lambda_n$, because it is hard to tell commata from semicolons, particularly when there are none, as is the case when there is only one row or column.) The notation $[\mathbf{v}_1,\ldots,\mathbf{v}_n]$ is also, usually without commata, the column notation for matrices employed in popular texts, like [La] or [Le]. Possibly because of the standard way, $\sum_j a_j\mathbf{v}_j$, of writing the linear combination of the vectors $\mathbf{v}_1,\ldots,\mathbf{v}_n$ with weights $a_1,\ldots,a_n$, some areas (e.g., wavelets and CAGD) have adopted the short-hand $aV = a[\mathbf{v}_1;\ldots;\mathbf{v}_n]$ for that weighted sum, and this is fine in isolation, but becomes awkward when the map $a \mapsto aV$ is to be composed with other linear maps, as maps are customarily written on the left of their argument.

To be sure, [Le] uses the notation $[\mathbf{v}_1,\ldots,\mathbf{v}_n]$ for what is called there an *ordered* basis and just begs to be called what it is, an (invertible) column map. The notation $\{\mathbf{v}_1;\ldots;\mathbf{v}_n\}$, used in some other standard text for an 'ordered basis', seems less helpful today, given `MATLAB`'s convention to use semicolons for vertical listings, and given that braces are usually reserved for sets. For this last reason, the notation $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ for a basis used in many texts (and even referred to there as a *set*) is likely to lead to confusion. In contrast, already [SB] denotes bases by $(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ (stressing the fact that a basis is a sequence, not a set) and, well before `MATLAB`, uses $[\mathbf{v}_1,\ldots,\mathbf{v}_n]$ for the matrix with columns $\mathbf{v}_1,\ldots,\mathbf{v}_n$, and even writes $(\mathbf{v}_1,\ldots,\mathbf{v}_n)a$ for the linear combination, with weight vector a, of the vectors $\mathbf{v}_j$ in an arbitrary vector space. From there, it is only a small step to employing the unifying notation $[\mathbf{v}_1,\ldots,\mathbf{v}_n]$ for the corresponding column map, and only one further small, but very helpful, step to thinking of bases as (invertible column) maps.

To be sure, it is at times much more convenient to index a basis by some index set I other than the set $\{1,2,\ldots,n\}$. In that case, the notation $[\mathbf{v}_i : i \in I]$ for the corresponding column map is perfectly handy. In fact, if the columns are pairwise distinct, one can even let the columns index themselves, i.e., consider, for a finite subset S of X, the corresponding column map $[S] : \mathbb{F}^S \to X : a \mapsto \sum_{\mathbf{v}\in S} a_\mathbf{v}\mathbf{v}$. It is in this way that those who prefer to think of a basis as a set would define the corresponding basis-map.

REFERENCES

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

[La] David C. Lay, *Applied Linear Algebra and its Applications, 2nd Edition*, Addison Wesley, Reading MA, 1997.

[Le] Steven J. Leon, *Linear Algebra with Applications, 4th Edition*, Prentice Hall, Englewood Cliffs NJ, 1990.

[P] Vlastimil Ptak, *Eine Bemerkung zur Jordanschen Normalform von Matrizen*, Acta Scientiarum Mathematicarum, Szeged, 17 (1956), pp. 190–194.

[SB] Hans Schneider and George Phillip Barker, *Matrices and Linear Algebra*, Holt, Rinehart and Winston, New York, 1968.