multivariate Hermite interpolation (talk at Guernavaca, 13apr99)

I am not really going to say something new, certainly not to the experts in the audience. Rather, I am going to try to persuade you that a certain point of view concerning interpolation might, at times, be very convenient. With that view in hand, I then consider Hermite interpolation, as a limit of Lagrange interpolation.

The particular point of view has its origin in elementary linear algebra, in the answer to the question: what is the inverse of a basis? leading to interpolation and convenient representation of interpolants, then to the limit of an interpolation process as the interpolation functionals approach a certain limit, and thence to Hermite interpolation.

I start off this elementary talk with something truly basic, namely the notion of a basis of a vector space.

Here is the definition of a basis, as no doubt you all have learned it, possibly already many years ago, and as it can be found in most current linear algebra textbooks. Here is a sample:

Definition (Strang'76). *A basis for a vector space is a set of vectors having two properties as once:*
(1) It is linearly independent.
(2) It spans the space.

This definition is a bit strange for the following reason (and you can check this out with your favorite Linear Algebra text, too). When it comes time to define the terms 'linear independent' and 'spanning' needed here, the elements of the set in question are always enumerated. E.g., Strang (loc.cit) writes:

Given a set of vectors $v_1, \ldots, v_k$, we look at their linear combinations $c_1v_1 + c_2v_2 + \cdots + c_kv_k$.

Further, here is a pop quiz: Assuming that

$$v_1 = x = v_2 \neq 0,$$

is the set

$$\{v_1, v_2\}$$

linearly independent? (NO ☐ SI ☐)

So, all of you who voted NO, i.e., most of you, don't really view a basis as a set, but as a sequence. This is slowly being realized by the textbook writers. E.g., in Strang's 1993 'Introduction to Linear Algebra', the above definition still appears, but with just one change, namely 'set' is replaced by 'sequence', and I am happy to take credit for that change (but acknowledge that my task of persuasion was made easier because I could point to the fact that, e.g., Bourbaki defines a basis as 'une famille', i.e., an 'indexed set', with those two properties).

But I am still not satisfied. I am now working on some textbook writers to get them to use a definition of basis that explicitly acknowledges the sole purpose of a basis.

That purpose, I claim, is to *provide linear representations*. As you all know, if X is the vector space and $(v_1, \ldots, v_k)$ is the basis in question, then, for every $x \in X$, there is exactly one choice of the scalar sequence $c_1, \ldots, c_k$ so that

$$x = c_1v_1 + \cdots + c_kv_k.$$

The linear independence says that there is *at most* one such choice, the spanning property says that there is *at least* one such choice. But, if that is the purpose, why not come right out and say so from the start and save the students a lot of confusion?

So, let me start from scratch.

By and large, we cannot compute with vectors in a vector space, we can only compute with scalars, i.e., elements in the underlying field $\mathbb{F}$, typically the real or the complex numbers. Hence, in order to compute with vectors, we have to represent them by scalars.

Ideally, such a representation is *linear*, hence a linear map

$$V : \mathbb{F}^k \to X$$

from scalar k-sequences

$$\mathbb{F}^k := \{c := (c(1), \ldots, c(k)) : c(j) \in \mathbb{F}\}$$

to the vector space X in question.

Any such linear map is of the form

$$V : \mathbb{F}^k \to X : c \mapsto c(1)v_1 + \cdots + c(k)v_k$$

with

$$v_j := V\mathbf{i}_j$$

the image under V of the jth coordinate vector in $\mathbb{F}^k$:

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

Conversely, any k-sequence $(v_1, \ldots, v_k)$ in X gives rise to a linear map from $\mathbb{F}^k$ to X, by the prescription

$$\mathbb{F}^k \to X : c \mapsto v_1c(1) + \cdots + v_kc(k) =: [v_1, \ldots, v_k]c.$$

Note how I have written here the weights to the right of the vectors, to stress that we are mapping c to something and in order to motivate the abbreviation $[v_1, \ldots, v_k]$ for this map.

This abbreviation is standard in case also X here is a coordinate space. Indeed, if $X = \mathbb{F}^m$, then we are accustomed to write its elements v_j as one-column matrices,

$$v_j =: \begin{bmatrix} v_j(1) \\ \vdots \\ v_j(m) \end{bmatrix}.$$

Correspondingly, in that case

$$c(1)v_1 + \cdots + c(k)v_k = Vc,$$

with V the (m, k)-matrix

$$V := [v_1, \ldots, v_k] = \begin{bmatrix} v_1(1) & \cdots & v_k(1) \\ \vdots & \ldots & \vdots \\ v_1(m) & \cdots & v_k(m) \end{bmatrix}$$

that has v_j as its jth column, all j. It it therefore an easy generalization to define, for an arbitrary sequence $(v_1, \ldots, v_k)$ in an arbitrary vector space X over $\mathbb{F}$, the map

$$[v_1, \ldots, v_k] : \mathbb{F}^k \to X : c \mapsto v_1c(1) + \cdots + v_kc(k),$$

and call it the **(k-)column map**, with **columns** $v_1, \ldots, v_k$.

- The k-column maps (with columns in X) comprise the linear maps from $\mathbb{F}^k$ to X.
- The set of linear combinations of the sequence $(v_1, \ldots, v_k)$ is the range of the corresponding column map, $[v_1, \ldots, v_k]$.
- *For any linear map* $A : X \to Y$,
$$A[v_1, \ldots, v_k] = [Av_1, \ldots, Av_k].$$

Note: In some areas, e.g., wavelets and CAGD, it has unfortunately become customary to write

$$\sum_j c(j)v_j = [c(1), \ldots, c(k)] \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix},$$

and that is certainly ok in isolation. The problem comes when we now want to apply some linear map A to the sum, something that's naturally handled by the column map notation I am trying to persuade you to use.

Having in this way identified the linear maps from $\mathbb{F}^k$ to X with the k-column maps (and their range with the linear combinations of their columns), the rest is very simple:

We would like a unique and linear representation of every $x \in X$, i.e., we would like an *invertible* linear map from some $\mathbb{F}^k$ to X, and that is exactly what the column map corresponding to a basis provides:

$$[v_1 \ldots, v_k] \text{ is } \left\{ \begin{array}{l} \text{1-1} \\ \text{onto} \\ \text{invertible} \end{array} \right\} \iff (v_1 \ldots, v_k) \text{ is } \left\{ \begin{array}{l} \text{lin.indep.} \\ \text{spanning} \\ \text{a basis} \end{array} \right\}$$

It is puzzling to me why, with these simple map notions of 1-1, onto, invertible available, there was any need to make up the additional terms 'linear independent', 'spanning', 'basis'. I recognize that it is too late to change matters now. But I do think we owe it to our students to stress this sole purpose of a basis, namely to provide a unique linear representation in terms of scalars, i.e., to provide the corresponding basis map, and so make it easier for them to handle those conventional terms.

Here is a silly example. You learn that every finite-dimensional vector space has a basis. So, what about the trivial space, $\{0\}$? We need an invertible linear map to $\{0\}$ from $\mathbb{F}^k$ for some k. Well, then $\mathbb{F}^k$ better have just one element, and that is exactly the case for $k = 0$ in which case $\mathbb{F}^k$ consists of all sequences with 0 entries. There is just such sequence, the empty sequence, (), and the basis map is the unique map from $\mathbb{F}^0$ to $\{0\}$, and is obviously invertible. It's the column map with no columns,

$$[\,]$$

(cf. `matlab`).

A more serious example (at least for my undergraduate students) is change of basis. Once we think in terms of the basis map, this problem is trivial:

If we know that $x = Wd$ for some basis(map) W, and V is any basis(map), then

$$x = VV^{-1}Wd = Vc, \quad \text{for } c := V^{-1}W.$$

In other words, $V^{-1}W$ is the so-called **transition matrix**, evidently a (square) matrix since it is a linear map from $\mathbb{F}^k$ to itself.

The same ease is experienced when wanting to represent the linear map $A : Y \to X$ with respect to bases W for Y and V for X, something we have to do if we wanted to actually work with the linear map A: If we know that $y = Wd$, then we compute $Ay = VV^{-1}AWd$, i.e.,

$$Ay = Vc, \quad \text{for } c := (V^{-1}AW)d.$$

Hence $V^{-1}AW$ is the needed matrix.

Of course, all of this you already know. So, next comes the question that I usually get at this point from one of the more active students: *how do I get hold of* V^{-1}, i.e.,

what is the inverse of a basis?

This is a serious question, particularly for students since, for them a map makes sense only if they the have a *formula* for it.

If X happens to be a coordinate space, i.e., necessarily $X = \mathbb{F}^k$, then $V = [v_1, \ldots, v_k]$ is just a square matrix and, correspondingly, V^{-1} is just its matrix inverse. In any other case, who knows???

It turns out that, in the general case, there is essentially just one recipe. You will recall special cases of this if not the recipe itself as soon as you see it. But I think it very worthwhile to stress this simple general recipe. In fact, I would judge my talk a success if all you took away from it is this recipe.

Recipe. *If $V : \mathbb{F}^k \to X$ is a basis(map) for the linear subspace X of the vector space Z, and $\Lambda^{\mathrm{t}} : Z \to \mathbb{F}^k$ is a linear map whose restriction to X is 1-1 or onto, then the matrix $\Lambda^{\mathrm{t}}V$ is invertible, and*

$$V^{-1} = (\Lambda^{\mathrm{t}}V)^{-1}\Lambda^{\mathrm{t}} \qquad \text{on} \quad X.$$

There's nothing to the proof: Since V is 1-1 and onto, while the restriction of Λ^{t} to $X = \operatorname{ran} V$ is 1-1 or onto, the matrix $\Lambda^{\mathrm{t}}V$ is 1-1 or onto, hence invertible (since it is square). Therefore

$$W := (\Lambda^{\mathrm{t}}V)^{-1}\Lambda^{\mathrm{t}}|_X$$

is well-defined, and $WV = \mathrm{id}$ by inspection, hence W is a left inverse for V, hence the inverse (since V is invertible, by assumption). □

One way to view this recipe is as the result of having discretized the *abstract* equation

$$V? = x$$

by applying Λ^{t} to both sides, thereby obtaining the *numerical* equation

$$\Lambda^{\mathrm{t}} V? = \Lambda^{\mathrm{t}} x.$$

For this to work, Λ^{t} better be 1-1 on X and $\Lambda^{\mathrm{t}} V$ better be square.

It follows that

$$V(\Lambda^{\mathrm{t}} V)^{-1}\Lambda^{\mathrm{t}}$$

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

$$x = V(\Lambda^{\mathrm{t}} V)^{-1}\Lambda^{\mathrm{t}} x$$

is the irredundant representation of $x \in X$ provided by the basis(map) V.

In practice, it may not be easy to know *a priori* whether $\Lambda^{\mathrm{t}}|_X$ is 1-1 or onto but one can often verify directly that $\Lambda^{\mathrm{t}} V$ is 1-1 or onto and then one even knows that, necessarily, V is a basis for its range.

Example

If Z is itself a coordinate space, $Z = \mathbb{F}^m$, say, hence V is an $m \times k$-matrix, then $\Lambda^{\mathrm{t}} : \mathbb{F}^k \to \mathbb{F}^m$ is, in effect, a $k \times m$-matrix. A standard choice for Λ^{t} is V^*, the (conjugate) transposed of V since, with V a basis for its range, i.e., V 1-1, so is V^*V and, being square, it is therefore invertible. The right side of the resulting formula

$$V^{-1} = (V^*V)^{-1}V^* \quad \text{on} \quad X = \operatorname{ran} V$$

is, of course, the **generalized inverse** of V. Correspondingly,

$$V(V^*V)^{-1}V^*$$

is the ortho-projector from $\mathbb{F}^m$ onto X. □

More generally, for any $\Lambda^{\mathrm{t}} : Z \to \mathbb{F}^k$ for which $\Lambda^{\mathrm{t}} V$ is invertible,

$$P := V(\Lambda^{\mathrm{t}} V)^{-1}\Lambda^{\mathrm{t}}$$

is the identity on its range, X, hence is a linear projector on Z to X. In fact, any linear projector from Z to X arises in this way.

This particular one is the unique linear projector to X for which

$$\Lambda^{\mathrm{t}} P z = \Lambda^{\mathrm{t}} z, \qquad \forall z \in Z,$$

i.e., for which Pz is the unique element in X that matches the information about z contained in the vector $\Lambda^{\mathrm{t}} z$. For this reason, Λ^{t} is called a *data map*, and P is an *interpolation scheme* for interpolation to the data supplied by Λ^{t}.

Example

Here is another standard example. I'll take

$$Z := \Pi,$$

the space of all (real-valued univariate) polynomials, as functions on the real line, and take

$$X := \Pi_n := \operatorname{ran} V, \qquad \text{with} \quad V := [()^j : j = 0{:}n],$$

with

$$()^j : t \mapsto t^j$$

my poor attempt at filling a painful hole in the notations provided by Mathematics. In other words, Π_n is, by definition, the linear space of all polynomials of degree $\le n$.

I'll take for Λ^{t} the restriction to a $(n+1)$-set $\mathrm{T} = \{\tau_0, \ldots, \tau_n\}$ of real numbers, so

$$\Lambda^{\mathrm{t}} = \Lambda^{\mathrm{t}}_{\mathrm{T}} :=: g \mapsto g|_{\mathrm{T}} = (g(\tau_i) : i = 0{:}n) \in \mathbb{R}^{n+1}.$$

How do I know that $\Lambda^{\mathrm{t}}_{\mathrm{T}} V$ is invertible? Since you all have learned this fact in some basic numerical analysis course, here is an opportunity to get more comfortable with this map point of view:

Recall the Lagrange polynomial:

$$\ell_i : t \mapsto \prod_{j \ne i} \frac{t - \tau_j}{\tau_i - \tau_j}$$

For it

$$\Lambda^{\mathrm{t}}_{\mathrm{T}} \ell_i = \mathbf{i}_i := (\delta_{ij} : j = 0{:}n).$$

Hence, for the corresponding column map

$$W := [\ell_0, \ldots, \ell_n],$$

we compute

$$\Lambda^{\mathrm{t}}_{\mathrm{T}} W = \Lambda^{\mathrm{t}} [\ell_i : i = 0{:}n] = [\Lambda^{\mathrm{t}} \ell_i : i = 0{:}n] = \mathrm{id}.$$

This implies that the $n+1$-column map W is 1-1, and since it is into Π_n, which is the range of the $n+1$-column map V, both W and V must be a basis(map) for Π_n. But then, also $\Lambda^{\mathrm{t}}_{\mathrm{T}} V$ must be invertible, since it is square and, e.g.,

$$\mathrm{id} = \Lambda^{\mathrm{t}}_{\mathrm{T}} W = \Lambda^{\mathrm{t}} V (V^{-1} W).$$

projectors, interpolation, change of basis

In the last example, the resulting map

$$P_{\mathrm{T}} = V (\Lambda^{\mathrm{t}}_{\mathrm{T}} V)^{-1} \Lambda^{\mathrm{t}}_{\mathrm{T}}$$

is the linear projector that associates $g \in Z$ with the unique polynomial of degree $\le n$ that matches g at the $(n+1)$-point set $\mathrm{T} = \{\tau_i : i = 0{:}n\}$, i.e., $P_{\mathrm{T}} g$ is the corresponding polynomial interpolant to g.

In full generality,

$$P g = V (\Lambda^{\mathrm{t}} V)^{-1} \Lambda^{\mathrm{t}} g$$

is the unique element of $\operatorname{ran} P = \operatorname{ran} V$ that matches g 'at' Λ^{t}, i.e., for which $\Lambda^{\mathrm{t}} P g = \Lambda^{\mathrm{t}} g$. It is in this sense that the general recipe for the inverse of a basis is, at the same time, the general recipe for interpolation from $\operatorname{ran} V = X$ to some data supplied by the map Λ^{t}.

In this sense, we can also think of it as nothing more than a *change of basis*. For, if $\Lambda^{\mathrm{t}} V$ is invertible, then so is $\Lambda^{\mathrm{t}}|_X$, and its inverse,

$$W := (\Lambda^{\mathrm{t}}|_X)^{-1},$$

is necessarily an invertible linear map from $\mathbb{F}^k$ to X, hence a basis for X. Knowing $\Lambda^{\mathrm{t}} x$ for some $x \in X$ means knowing nothing more than knowing the coordinates of x with respect to the basis W, from which our formula $Px = V(\Lambda^{\mathrm{t}} V)^{-1} \Lambda^{\mathrm{t}} x$ merely constructs the coordinates $(\Lambda^{\mathrm{t}} V)^{-1} \Lambda^{\mathrm{t}} x$ of x with respect to V.

analysis and synthesis; row maps and column maps

The maps V and Λ^{t} play dual roles in this discussion.

$\Lambda^{\mathrm{t}} : Z \to \mathbb{F}^k$ plays the role of a **data map**, or an **analysis operator**, as it *extracts* numerical information from the elements of Z.

$V : \mathbb{F}^k \to X$ goes the other way; from such numerical information, V (re)*constructs* an element of X, it is a **synthesis operator**.

Any such V is characterized by its columns. Dually, every $\Lambda^{\mathrm{t}} : Z \to \mathbb{F}^k$ is characterized by the linear functionals

$$\lambda_i : Z \to \mathbb{F} : z \mapsto (\Lambda^{\mathrm{t}} z)(i), \qquad i = 1{:}k,$$

i.e., by the linear functionals defined by

$$\Lambda^{\mathrm{t}} z =: (\lambda_i z : i = 1{:}k), \qquad z \in Z.$$

That being so, and in view of the special case $Z = \mathbb{F}^m$, it is natural to write such Λ^{t}, more explicitly, as

$$\Lambda^{\mathrm{t}} = [\lambda_1, \ldots, \lambda_k]^{\mathrm{t}},$$

calling it a **(k-)row map**, with **rows** $\lambda_1, \ldots, \lambda_k$.

E.g., for polynomial interpolation at the points of T,

$$\Lambda^{\mathrm{t}} = \Lambda^{\mathrm{t}}_{\mathrm{T}} = |_{\mathrm{T}} = [\delta_{\tau_i} : i = 0{:}n]^{\mathrm{t}},$$

with

$$\delta_\tau : g \mapsto g(\tau)$$

the linear functional of point evaluation at τ.

In this way, the matrix

$$\Lambda^{\mathrm{t}} V = [\lambda_1, \ldots, \lambda_k]^{\mathrm{t}} [v_1, \ldots, v_k] = (\lambda_i v_j : i, j = 1{:}k)$$

is the **Gramian** of the two sequences, while

$$V \Lambda^{\mathrm{t}} = [v_1, \ldots, v_k][\lambda_1, \ldots, \lambda_k]^{\mathrm{t}} : z \mapsto \sum_j (\lambda_j z) v_j$$

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

simplify the formula $P = V(\Lambda^{\mathrm{t}} V)^{-1} \Lambda^{\mathrm{t}}$

In particular, V and Λ^{t} might be **dual to each other**, i.e.,

$$\Lambda^{\mathrm{t}} V = \mathrm{id},$$

in which case the formula $P = V(\Lambda^{\mathrm{t}} V)^{-1} \Lambda^{\mathrm{t}}$ simplifies, i.e., then

$$V \Lambda^{\mathrm{t}}$$

is a linear projector onto $\operatorname{ran} V$.

E.g., $V = [()^j : j = 0{:}n]$ is dual to

$$\Lambda^{\mathrm{t}} : z \mapsto (D^j z(0)/j! : j = 0{:}n),$$

leading to the truncated Taylor expansion at 0:

$$z(t) \approx \sum_j t^j D^j z(0)/j!.$$

If the Gramian is merely invertible, then any factorization

$$\Lambda^{\mathrm{t}} V = AC$$

of the Gramian into square, hence invertible, matrices A and C leads to the modified maps

$$\widehat{V} := VC^{-1}, \qquad \widehat{\Lambda}^{\mathrm{t}} := A^{-1}\Lambda^{\mathrm{t}},$$

which are dual to each other but describe the same linear projector:

$$\widehat{V}\widehat{\Lambda} = P = V(\Lambda^{\mathrm{t}}V)^{-1}\Lambda^{\mathrm{t}}.$$

For our example of polynomial interpolation, a simple choice is

$$A = \mathrm{id}, \quad C = \Lambda^{\mathrm{t}}V,$$

hence

$$\widehat{\Lambda}^{\mathrm{t}} = \Lambda^{\mathrm{t}}$$

and therefore, necessarily,

$$\widehat{V} = V(\Lambda^{\mathrm{t}}V)^{-1} = [\ell_i : i = 0{:}n]$$

the column map whose columns are the relevant Lagrange polynomials. By expanding ℓ_i in powers, we get in this way an explicit formula for the inverse of the matrix $\Lambda^{\mathrm{t}}V$, the **Vandermonde** matrix, socalled by Lebesgue.

A more interesting factorization of the Vandermonde is the one obtained by Gauss elimination (without pivoting): Here

$$\Lambda^{\mathrm{t}}V = LU,$$

with L lower triangular and U upper triangular. Such a factorization if it exists is unique up to the choice of the diagonal elements of either L or U. I'll choose U to be *unit* upper triangular.

Then also U^{-1} is unit upper triangular, hence $\widehat{v}_j$, as column j of

$$[\ldots, \widehat{v}_j, \ldots] := \widehat{V} = VU^{-1} = [()^0, \ldots, ()^n] \begin{bmatrix} & \times & \\ & \vdots & \\ & \times & \\ \cdots & \mathbf{1} & \cdots \\ & 0 & \\ & \vdots & \\ & 0 & \end{bmatrix},$$

has leading term t^j, i.e., $\widehat{v}_j(t) = t^j + \mathbf{lot}$.

Further,

$$[\ldots, \Lambda^{\mathrm{t}}\widehat{v}_j, \ldots] = \Lambda^{\mathrm{t}}\widehat{V} = L\widehat{\Lambda}\widehat{V} = L = \begin{bmatrix} & 0 & \\ & \vdots & \\ \cdots & 0 & \cdots \\ & \times & \\ & \vdots & \\ & \times & \end{bmatrix}$$

is *lower* triangular, hence $\widehat{v}_j$ vanishes on $t_0, \ldots, t_{j-1}$. Hence, altogether,

$$\widehat{v}_j = \prod_{i<j}(\cdot - t_i), \qquad j = 0{:}n,$$

and we recognize these as the polynomials appearing in the Newton form for the interpolating polynomial.

Again, since L is lower triangular, $\widehat{\lambda}_i$, as row i of $L^{-1}\Lambda^{\mathrm{t}}$, is a linear combination of the evaluations $\delta_{\tau_0}, \ldots, \delta_{\tau_i}$, and, since

$$\begin{bmatrix} \cdots \\ \widehat{\lambda}_i v_1, \ldots, \widehat{\lambda}_i v_n \\ \cdots \end{bmatrix} = \widehat{\Lambda}V = \widehat{\Lambda}\widehat{V}U = U = \begin{bmatrix} & & \cdots & & \\ 0 & \cdots & 0 & \mathbf{1} & \cdots \\ & & \cdots & & \end{bmatrix}$$

is unit upper triangular, $\widehat{\lambda}_i$ vanishes on $()^0, \ldots, ()^{i-1}$ and has the value 1 on $()^i$, hence

$$\widehat{\lambda}_i = \boldsymbol{\Delta}(\tau_0, \ldots, \tau_i),$$

the divided difference at the point sequence $(\tau_0, \ldots, \tau_i)$ (in Velvel Kahan's felicitous notation).

I like the idea that, at least in the univariate context, divided differences arise naturally from Gauss elimination, as this points to a natural generalization of divided differences to the multivariate context.

You may have fun working out the details of the following related claim:

. *The Gram-Schmidt algorithm of orthogonalization of a linearly independent sequence* $(v_1, \ldots, v_k)$ *is Gauss Elimination without pivoting applied to the Gram matrix*

$$V^*V = (\langle v_j, v_i \rangle : i, j = 1{:}k).$$

Hermite is coalescence

So far, this has been a very leisurely walk through very familiar territory but, perhaps, with very strange glasses on. My claim is that these glasses are very useful once you get used to them. I mainly took the time for this leisurely walk in order to give you a chance to get used to them.

I am now ready to take on the stated topic of this talk, namely Hermite interpolation.

To me, it is a question of the dependence of, e.g., polynomial interpolation

$$P_{\mathrm{T}} = V_{\mathrm{T}} \Lambda_{\mathrm{T}}^{\mathrm{t}}, \qquad V_{\mathrm{T}} := [\ell_j : j = 0{:}n], \;\; \Lambda_{\mathrm{T}}^{\mathrm{t}} := |_{\mathrm{T}},$$

on the sequence $\mathrm{T} = (\tau_i : i = 0{:}n)$. Specifically, I want to know what happens as $\mathrm{T} \to \Sigma$ with some of the entries of Σ coincident. (Of course, you all know what happens, so this is still one more bit of preparation for the multivariate case.)

If we look directly at V_{T} and $\Lambda_{\mathrm{T}}^{\mathrm{t}}$, we notice that $\Lambda_{\mathrm{T}}^{\mathrm{t}} \to \Lambda_{\Sigma}^{\mathrm{t}}$, while V_{T} fails to converge. Does this mean that P_{T} fails to converge? Of course not. It only means that this particular description of P_{T} isn't very helpful here.

Let's look again at our general description

$$P = V(\Lambda^{\mathrm{t}} V)^{-1} \Lambda^{\mathrm{t}}$$

of a linear projector. For V here, we can take any basis of $\operatorname{ran} P$. What is our freedom as regards the data map Λ^{t}?

By going to the dual, we see that

$$P' = \Lambda((\Lambda^{\mathrm{t}} V)^{-1})' V',$$

hence that

$$(\Lambda^{\mathrm{t}})' = \Lambda = [\lambda_1, \ldots, \lambda_k]$$

can be chosen as any particular basis for $\operatorname{ran} P'$. While $\operatorname{ran} P$ provides the interpolant, $\operatorname{ran} P'$ provides the **interpolation functionals**, i.e.,

$$\operatorname{ran} P' = \operatorname{ran} \Lambda = \{\lambda \in \Pi' : \lambda = \lambda P\},$$

a description of the information to be matched.

With this, our task is simple: check what happens to $\operatorname{ran} P_{\mathrm{T}}$ and $\operatorname{ran} P'_{\mathrm{T}}$ as $\mathrm{T} \to \Sigma$.

Since $\operatorname{ran} P_{\mathrm{T}} = \Pi_n$ regardless of T, nothing happens there. What about $\operatorname{ran} P'_{\mathrm{T}} = \operatorname{ran}[\delta_\tau : \tau \in \mathrm{T}]$?

Well, you all know the answer, as one of the payoffs of the Newton form. Extend T in any way whatsoever to a bounded infinite sequence $\tilde{\mathrm{T}}$ and extend Σ by the same terms. Then still $\tilde{\mathrm{T}} \to \tilde{\Sigma}$, and

$$\widehat{v}_j := \prod_{i<j} (\cdot - \tau_i), \qquad j = 0, 1, \ldots$$

provides a basis for Π, and this basis converges elementwise to

$$\widehat{w}_j := \prod_{i<j}(\cdot - \sigma_i), \qquad j = 0, 1, \ldots$$

as $\tilde{\mathrm{T}} \to \tilde{\Sigma}$. Since

$$\widehat{\lambda}_i \widehat{v}_j = \delta_{ij},$$

it follows that

$$\widehat{\lambda}_i = \mathbf{\Delta}(\tau_0, \ldots, \tau_i)$$

converges to the corresponding coordinate functional for the basis

$$\widehat{W} := [\widehat{w}_j : j = 0, 1, \ldots],$$

usually also denoted

$$\mathbf{\Delta}(\sigma_0, \ldots, \sigma_i).$$

Hence

$$P_{\mathrm{T}} \to P_\Sigma := \widehat{V}_\Sigma \widehat{\Lambda}_\Sigma^{\mathrm{t}}$$

with

$$\widehat{\Lambda}_\Sigma := [\mathbf{\Delta}(\sigma_0, \ldots, \sigma_i) : i = 0{:}n]$$

1-1, hence a basis for its range.

What exactly is its range? Since this range is finite-dimensional, we know that

$$\operatorname{ran} \widehat{\Lambda}_\Sigma = \{\lambda \in \Pi' : \ker \lambda \supset \ker \widehat{\Lambda}_\Sigma^{\mathrm{t}}\}.$$

Since

$$\ker \widehat{\Lambda}_\Sigma^{\mathrm{t}} = (\cdot - \sigma_0) \cdots (\cdot - \sigma_n)\Pi,$$

this implies that $\operatorname{ran} \widehat{\Lambda}_\Sigma$ contains the sequence

$$(\delta_z D^j : 0 \le j < \#\{i : z = \sigma_i\})$$

which is linearly independent and of length $n+1$, hence necessarily a basis for the space of 'interpolation functionals' for the limiting P_Σ.

multivariate

Finally, I am ready to discuss the multivariate situation. Now

$$\mathrm{T}$$

is a sequence in $\mathbb{R}^d$ for some $d > 1$ and the question is again:

$$\lim_{\mathrm{T} \to \Sigma} \operatorname{ran}[\delta_\tau : \tau \in \mathrm{T}] = ???$$

As you surely already know, at least from Mariano Gasca's talk yesterday, we don't have any divided difference to help us since it isn't at all obvious how we should choose an interpolating polynomial. Yet, even without any particular $\operatorname{ran} P$ in mind, it makes sense to consider linear projectors P with data map $\Lambda_{\mathrm{T}}^{\mathrm{t}}$ and to wonder what happens to the corresponding space of interpolation functionals as $\mathrm{T} \to \Sigma$.

So, let's explore, by taking

$$\mathrm{T} = (\tau_0, \ldots, \tau_n)$$

with $n = 1$.

If, e.g.,

$$\tau_0 = \vartheta, \quad \tau_1 = \vartheta + \varepsilon\xi,$$

then, as $\varepsilon \to 0$, we get
$$\lim_{\varepsilon\to 0}(\delta_{\tau_1} - \delta_{\tau_0})/\varepsilon = \lim_{\varepsilon\to 0}(\delta_{\vartheta+\varepsilon\xi} - \delta_\vartheta)/\varepsilon = \delta_\vartheta D_\xi$$
i.e., the value at ϑ of the **directional derivative**
$$D_\xi := \sum_{i=1}^{d} \xi(i) D_i.$$

If, on the other hand,
$$\tau_1 = \vartheta + \varepsilon(\cos(1/\varepsilon), \sin(1/\varepsilon)),$$
then $\operatorname{ran}[\tau_0, \tau_1]$ fails to have a limit as $\varepsilon \to 0$.

So we must be prepared to be modest in our expectations.

We consider first the simplest possible situation, namely that
$$\mathrm{T} = \varepsilon\Xi, \qquad \varepsilon \to 0.$$

Let
$$\lambda := \sum_{\xi\in\Xi} c(\xi)\delta_\xi,$$
and consider
$$\lambda_\varepsilon p := \left(\sum_{\xi\in\Xi} c(\xi)\delta_{\varepsilon\xi}\right) p = \lambda p(\varepsilon\cdot)$$
for some polynomial p as $\varepsilon \to 0$. Write the polynomial in power form, i.e.,
$$p = \sum_\alpha ()^\alpha \widehat{p}(\alpha),$$
with
$$()^\alpha : \mathbb{R}^d \to \mathbb{R} : t \mapsto t(1)^{\alpha(1)} \cdots t(d)^{\alpha(d)},$$
and
$$\widehat{p}(\alpha) := D^\alpha p(0)/\alpha!, \qquad \alpha \in \mathbb{Z}_+^d$$
its power coefficients, and
$$D^\alpha := D_1^{\alpha(1)} \cdots D_d^{\alpha(d)}$$
the corresponding partial derivative, and
$$\alpha! := \alpha(1)! \cdots \alpha(d)!.$$
We'll also use the standard notation
$$p(D) := \sum_\alpha \widehat{p}(\alpha) D^\alpha$$
for the corresponding constant coefficient differential operator, as well as the standard notation
$$|\alpha| := \alpha(1) + \cdots + \alpha(d).$$

With these notations, we compute
$$\begin{aligned}
\lambda_\varepsilon p = \lambda p(\varepsilon\cdot) &= \sum_{\xi\in\Xi} c(\xi) \sum_\alpha (\varepsilon\xi)^\alpha \widehat{p}(\alpha) \\
&= \sum_j \varepsilon^j \sum_{|\alpha|=j} \underbrace{\sum_{\xi\in\Xi} c(\xi)\xi^\alpha}_{\lambda()^\alpha} \widehat{p}(\alpha) \\
&= \sum_{j\ge \operatorname{order}\lambda} \varepsilon^j \sum_{|\alpha|=j} \lambda()^\alpha \widehat{p}(\alpha)
\end{aligned}$$

with

$$\operatorname{order}\lambda := \min\{|\alpha| : \lambda()^\alpha \neq 0\}.$$

Therefore

$$\begin{aligned}\lim_{\varepsilon\to 0} \lambda_\varepsilon p/\varepsilon^{\operatorname{order}\lambda} = \sum_{|\alpha|=\operatorname{order}\lambda} \lambda()^\alpha \widehat{p}(\alpha) &= \sum_{|\alpha|=\operatorname{order}\lambda} \lambda()^\alpha \frac{1}{\alpha!} D^\alpha p(0) \\ &= q(D)p(0),\end{aligned}$$

with

$$q := \sum_{|\alpha|=\operatorname{order}\lambda} \sum_{\xi\in\Xi} c(\xi)\frac{\xi^\alpha}{\alpha!}\,()^\alpha \;= ???$$

a certain polynomial, that we'll look further into. Note that, in the univariate case, this sum would only have one term in it and, correspondingly, the limit is just a scalar multiple of the $(\operatorname{order}\lambda)$-th derivative at the origin, just as expected. In the multivariate case, things are much more complicated.

Yet, as we look further into this polynomial q, we'll also discover real beauty.

What does the term $\xi^\alpha/\alpha!$ remind you of? The *exponential function*!

Right, you recall

$$\mathrm{e}_\xi : t \mapsto \mathrm{e}^{\xi\cdot t} = \sum_j (\xi\cdot t)^j/j! = \sum_\alpha \frac{\xi^\alpha}{\alpha!}\, t^\alpha,$$

the **exponential with frequency** ξ, with

$$\xi\cdot t := \sum_i \xi(i)t(i)$$

the standard scalar product in $\mathbb{R}^d$.

Define

$$f := \sum_{\xi\in\Xi} c(\xi)\mathrm{e}_\xi = \sum_j \underbrace{\sum_{|\alpha|=j}\sum_{\xi\in\Xi} c(\xi)\frac{\xi^\alpha}{\alpha!}\,()^\alpha}_{=:\, f^{[j]}}.$$

And there we discover q:

$$q = f^{[\operatorname{order}\lambda]}.$$

In other words: if we organize f into its *homogeneous terms*,

$$f = f^{[0]} + f^{[1]} + \cdots,$$

then we find that $f^{[\operatorname{order}\lambda]}$ is the first such term that is non-zero. For that reason, we call it the **least** or **initial** term of f, and denote it by

$$f_\downarrow.$$

Conclusion.

$$\lim_{\varepsilon\to 0} \operatorname{ran}[\delta_{\varepsilon\xi} : \xi\in\Xi] \;\supseteq\; \{\delta_0 q(D) : q \in \Pi_\Xi\}$$

with

$$\Pi_\Xi \;:=\; \{f_\downarrow : f \in \operatorname{ran}[\mathrm{e}_\xi : \xi\in\Xi]\}.$$

Claim.

$$\lim_{\varepsilon\to 0} \operatorname{ran}[\delta_{\varepsilon\xi} : \xi\in\Xi] = \{\delta_0 q(D) : q \in \Pi_\Xi\}$$

since, for any $\varepsilon > 0$,

$$\dim\operatorname{ran}[\delta_{\varepsilon\xi} : \xi\in\Xi] = \#\Xi = \dim\Pi_\Xi.$$

Let's try the first equality first. It is a special case of the fact that, for any finite T, $[\delta_\tau : \tau \in \mathrm{T}]$ is 1-1, or, equivalently,

$$\Lambda_{\mathrm{T}}^{\mathrm{t}} = [\delta_\tau : \tau \in \mathrm{T}]^{\mathrm{t}}$$

is onto. Is it obvious? Here is a quick proof: Define

$$w_\tau : t \mapsto \prod_{\sigma \in \mathrm{T}\backslash\tau} (t - \sigma)\cdot(\tau - \sigma)$$

and observe that

$$\forall\{\sigma \in \mathrm{T}\}\;\; w_\tau(\sigma) = 0 \quad \iff \quad \sigma \neq \tau.$$

Hence

$$\Lambda_{\mathrm{T}}^{\mathrm{t}}[w_\tau / w_\tau(\tau) : \tau \in \mathrm{T}] = \mathrm{id}.$$

Incidentally, look what we have just produced here: Since each

$$\ell_\tau := w_\tau / w_\tau(\tau)$$

is a polynomial of degree $< \#\mathrm{T}$,

$$[\ell_\tau : \tau \in \mathrm{T}]\Lambda_{\mathrm{T}}^{\mathrm{t}}$$

provides a polynomial interpolant to data at an arbitrary finite pointset T. The interpolant has some nice properties. For example, it is symmetric in the points in T, and, in the univariate case, it reduces to the standard polynomial interpolant. Remember how earlier I bemoaned the fact that there doesn't seem to be a 'natural' polynomial interpolant to data at an arbitrary pointset? So, why am I not happy with this one?

It has an unnecessarily high degree. E.g., for three generic points in the plane, it will give a quadratic interpolant while the 'natural' interpolant is linear.

Back to our problem of identifying

$$\Pi_\Xi \;=\; \{f_\downarrow : f \in \operatorname{ran}[\mathrm{e}_\xi : \xi \in \Xi]\}.$$

We know that it is in the limit of a sequence of linear spaces, each of dimension $\#\Xi$. Consequently, we now know that

$$\dim \Pi_\Xi \le \#\Xi.$$

To get equality, hence get the second needed equality in the (second display of the) Claim, we merely have to find a linearly independent sequence of $\#\Xi$ elements in Π_Ξ.

Let's start with the basic question: how do we get *any* element of Π_Ξ?

If we take $f(t) = \mathrm{e}_\xi(t) = \sum_j (\xi\cdot t)^j / j!$, then, obviously,

$$\mathrm{e}_{\xi\downarrow} = 1.$$

So far, so good. If $\#\Xi = 1$, that is it. Otherwise, there is $\zeta \in \Xi\backslash\xi$, and we could take

$$f := \mathrm{e}_\zeta - \mathrm{e}_\xi = (\zeta - \xi)\cdot + \mathbf{hot},$$

giving us

$$f_\downarrow(t) = (\zeta - \xi)\cdot t,$$

a linear polynomial.

If there are more points in Ξ, we can now form more complicated linear combinations

$$f = \sum_{\xi \in \Xi} c(\xi)\mathrm{e}_\xi = \sum_\alpha ()^\alpha \widehat{f}(\alpha)$$

with the goal of having the first so many power coefficients $\widehat{f}(\alpha)$ equal zero.

The first so many??? Well, choose any ordering of the multi-index set $\mathbb{Z}_+^d$ you like as long as it is compatible with degree, i.e., as long as

$$|\alpha| < |\beta| \quad \Longrightarrow \quad \alpha < \beta.$$

(E.g., the lexicographic ordering will do.) With that, we can think of

$$\widehat{f} = (\widehat{f}(\alpha) : \alpha \in \mathbb{Z}_+^d)$$

as a(n infinite) vector. Since $\widehat{f}$ is a weighted sum of the coefficient vectors

$$\widehat{\mathrm{e}_\xi} := (\xi^\alpha/\alpha! : \alpha \in \mathbb{Z}_+^d)$$

of the functions e_ξ, all $\xi \in \Xi$, we are, in effect, trying to form linear combinations of the rows of the matrix

$$G := (\xi^\alpha/\alpha! : \xi \in \Xi, \alpha \in \mathbb{Z}_+^d) \;=\; \begin{bmatrix} & \cdots & \\ \cdots & \xi^\alpha/\alpha! & \cdots \\ & \cdots & \end{bmatrix}$$

in such a way that the first so many entries are zero.

Is that an operation you have come across before???

Of course, that is exactly what Gauss elimination is designed to do!

Gauss elimination produces the factorization

$$G = LU,$$

with L invertible and with U in row-echelon form. Since we know that G is onto (i.e., the rows of G are linearly independent (how??)), we know that each row of U is nontrivial. Since L is invertible, the rows of U are weighted sums of the rows of G, hence each row of U provides us with an element of Π_Ξ. Moreover, since U is in row-echelon form, the resulting sequence of $\#\Xi$ elements of Π_Ξ is linearly independent, hence necessarily a basis for Π_Ξ. This finishes the proof of the above Claim.

why e_ξ here?

Next, here is an explanation for the still unexplained appearance of the exponential functions e_ξ.

Consider the pairing

$$A_0 \times \Pi : (g, p) \to \langle g, p\rangle := \sum_\alpha \widehat{g}(\alpha)\,\alpha!\,\widehat{p}(\alpha),$$

in which p is a polynomial, hence the sum has only finitely many nonzero terms regardless of the choice of the sequence $\widehat{g}$, i.e., the pairing makes sense for an arbitrary **formal power series**

$$g = \sum_\alpha ()^\alpha \widehat{g}(\alpha).$$

Try, in particular,

$$g = \mathrm{e}_\xi = \sum_\alpha ()^\alpha \xi^\alpha/\alpha!,$$

to get

$$\langle \mathrm{e}_\xi, p\rangle = \sum_\alpha \xi^\alpha \widehat{p}(\alpha) = p(\xi).$$

Representation. e_ξ *represents* δ_ξ *on* Π *wrto the given pairing.*

Now also the appearance of Π_Ξ makes more sense. In effect,

$$\Pi_\Xi = \lim_{\varepsilon\to 0} \mathrm{ran}[\mathrm{e}_{\varepsilon\xi} : \xi \in \Xi]$$

In other words, Π_Ξ is the limiting polynomial space of the representers of the interpolation functionals $\mathrm{ran}[\delta_{\varepsilon\xi} : \xi \in \Xi]$ as $\varepsilon \to 0$.

To be sure, earlier we wrote those limiting linear functionals as

$$p \to q(D)p(0), \qquad q \in \Pi_\Xi.$$

So, here is a final observation, easy to verify:

$$\begin{aligned} q(D)p(0) &= \sum_\alpha D^\alpha q(0) \frac{1}{\alpha!} D^\alpha p(0) \\ &= \sum_\alpha \widehat{q}(\alpha)\, \alpha!\, \widehat{p}(\alpha) \;=\; \langle q, p\rangle. \end{aligned}$$

the general case

If $\mathrm{T} \to \Sigma$ with Θ the distinct elements in Σ, then, in the nicest situation,

$$\mathrm{T} \sim (\vartheta + \varepsilon\Xi_\vartheta : \vartheta \in \Theta), \qquad \varepsilon \to 0,$$

for certain point sets Ξ_ϑ.

Since

$$\mathrm{e}_{\vartheta+\varepsilon\xi} = \mathrm{e}_\vartheta \mathrm{e}_{\varepsilon\xi},$$

the corresponding space of representers is

$$\sum_{\vartheta\in\Theta} \mathrm{e}_\vartheta\, \mathrm{ran}[\mathrm{e}_{\varepsilon\xi} : \xi \in \Xi_\vartheta]$$

and this converges, as $\varepsilon \to 0$, to

$$\sum_{\vartheta\in\Theta} \mathrm{e}_\vartheta \Pi_{\Xi_\vartheta}$$

leading to the limiting interpolation functionals

$$\sum_{\vartheta\in\Theta} \{\delta_\vartheta q(D) : q \in \Pi_{\Xi_\vartheta}\}.$$

properties of Π_Ξ

It turns out that Π_Ξ has many remarkable properties. I mention only a few:

Properties of Π_Ξ.

- Π_Ξ *is* **dilation-invariant**, *i.e.*,

$$q \in \Pi_\Xi,\ r > 0 \quad \Longrightarrow \quad q(r\cdot) \in \Pi_\Xi.$$

- Π_Ξ *is* **translation-invariant**, *i.e.*,

$$q \in \Pi_\Xi, \tau \in \mathbb{R}^d \quad \Longrightarrow \quad q(\cdot - \tau) \in \Pi_\Xi,$$

  hence Π_Ξ *is* D**-invariant**, *i.e.*,

$$q \in \Pi_\Xi, \alpha \in \mathbb{Z}_+^d \quad \Longrightarrow \quad D^\alpha q \in \Pi_\Xi.$$

  Both of these follow from the following perhaps most intriguing property,

-

$$\Pi_\Xi = \cap_{p|_\Xi = 0} \ker p^{[\deg p]}(D).$$

Conjecture. *Every dilation-invariant and translation-invariant (finite-dimensional) polynomial space is necessarily of the form* Π_Ξ *for some set* Ξ*.[added 20apr08: true for* $d = 2$*, false for* $d > 2$ *(B. Shekhtman)]*

This motivates the following definition:

Definition: Hermite interpolation. *(Multivariate) Hermite interpolation occurs when the space of interpolation functionals is of the form*

$$\operatorname{ran} P' = \sum_{\vartheta\in\Theta} \{\delta_\vartheta q(D) : q \in Q_\vartheta\}$$

for some finite point set Θ and some finite-dimensional dilation- and translation-invariant polynomial spaces Q_ϑ, $\vartheta \in \Theta$. [added 20apr08: Have, meanwhile, restricted the term "Hermite interpolation" to mean the limits of Lagrange interpolation, in view of Shekhtman's proof that, for $d > 2$, not every ideal interpolation is Hermite interpolation in this restricted sense.]

The important property here is the translation invariance of the Q_ϑ. For this is a necessary and sufficient condition for P to be an **ideal interpolation scheme**, as defined by Garrett Birkhoff. This means that

$$(\operatorname{ran} P')_\perp = \ker P$$

is an **ideal**, i.e., closed under pointwise multiplication by any polynomial. By Hilbert's Nullstellensatz, this guarantees that there is a *finite* set H of polynomials, so that

$$\operatorname{ran}(\mathrm{id} - P) = \ker P = \sum_{h\in H} h\Pi,$$

hence gives hope that, eventually, one may have in hand reasonable error formulae for the error

$$g - Pg = (\mathrm{id} - P)g$$

in the interpolant Pg to given g.

But that is another whole, well, quite unfinished, story.

References

The first part of this talk is based on an unpublished paper, entitled 'The inverse of a basis(map)' (10aug03: which, meanwhile, has appeared as 'What is the inverse of a basis?', BIT; 41(5); 2001; 880–890; and) which, in turn, is based on the paper

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

The second part is based on joint work with Amos Ron concerning the 'least polynomial interpolant' at an arbitrary set Ξ of data sites. Check

`www.cs.wisc.edu/~deboor/multiint.html/`

That work, in turn, is based on

[2] N. Dyn and A. Ron, *On multivariate polynomial interpolation*, Algorithms for Approximation II (J. C. Mason and M. G. Cox, eds.), Chapman & Hall, London, 1990, pp. 177–184.

in which the above definition of Hermite interpolation already occurs.

The conjecture is from

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

which also discusses in detail the fact that Π_Ξ is the 'limit at the origin' of $\operatorname{ran}[\mathrm{e}_\xi : \xi \in \Xi]$; see also

[4] C. de Boor and A. Ron, *The limit at the origin of a smooth function space*, Approximation Theory VI (C. Chui, L. Schumaker, and J. Ward, eds.), Academic Press, New York, 1989, pp. 93–96.

Finally, Birkhoff's definition of an ideal interpolation scheme appeared in

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

C. de Boor
(updated 08feb01 to change Λ to Λ^{t})
(updated 10aug03 to update a reference)
(updated 20apr08 to correct misprints)

`ftp.cs.wisc.edu/Approx/encoan.pdf`