Jad Nohra, Private Note. Revised: September 10, 2016

Mathematical Rags

Elementary Covariance Through Uncanonical Vector Representation.

I Motivation

1. Parametrization is a topic of geometry. What about parametrization within finite dimensional linear (vector) spaces? It is not a topic because it is trivial. However, we think it does deserve a short look.

2. Let us work within a linear space V of dimension n, over $\mathcal{F}$. Given a chosen basis $B = \{b_i\}$, one chooses a standard bijective representation function

$$r_B : \mathcal{F}^n \to V,$$

such that it produces a point[1] $a \in V$ from a point $\underline{a} \in \mathcal{F}^n$ in the following manner:

$$a = \underline{a}_1 b_1 + \cdots + \underline{a}_n b_n.$$

Using a loose summation notation this is:

$$a = r(a_*) = \underline{a}_i b_i,$$

where (x_*) stands for $(x_1, \ldots, x_n)$.

3. This is a useful and obvious choice, but a choice nevertheless. Without going into details of multilinear algebra, one can take any map ϕ that preserves linear independence and meaningfulness of representation of a point with respect to a basis (in short, bijective multilinear), and still have a valid unique representation of a:

$$a = \underline{a}_i \phi_i(b_*), \tag{1}$$

4. We can view all such ϕ as 'parametrizations' of V, and Equation 1 as a 'uncanonical vector representation'. This allows us to consider not only choices of basis, but also choices of parametrization. This is worthwhile because it gives a peek at the concept 'variance' (i.e covariance and contravariance), already within the realm of elementary linear algebra, without explicitly talking about duality, linear functionals or tensors, indicating that in some sense, the concept is quite elementary.

Rag: A large roofing slate that is rough on one side.

[1] We say point instead of vector to put ourselves into a more geometric mindset for the reason we explained.

II Variance

5. In the abstract setting, one considers a point to be independent of its representation. Even though we forego the abstract treatment, we can still express this by remembering that any property that makes sense in the abstract setting can be mirrored in the coordinate setting.

6. Without repeating known details pertaining to change of basis in linear algebra, without mentioning choices of basis when they do not matter for our purpose, working directly with matrices, let us consider the following:

- A point a which we identify with its representation in some basis.
- A basis B which we identify with a matrix, such that

$$a = B\underline{a}_i.$$

- Another basis C, such that

$$\begin{aligned} a &= C\underline{\alpha}_i, \\ C &= BL, \end{aligned}$$

  L being the change-of-basis matrix.

7. Our coordinate based version of representation independence requires the following:

$$Ia = B\underline{a}_i = C\underline{\alpha}_i,$$

we have:

$$\begin{cases} \underline{\alpha_i} & = L^{-1}\underline{a_i} \\ C & = BL \end{cases} \tag{2}$$

8. We can see that coordinates and bases change in 'opposite' ways. The first is called contravariant and the second covariant. In what follows, we make use of what we called paramterization in Equation 1 to show that one can choose it such that something different than the above happens.

9. Let us denote our parametrization function as F, a function acting multilinearly on a basis (its matrix). Instead of searching for a useful parametrization, we emulate the construction of dual basis and chose:

$$F : X \mapsto X^{-T},$$

that is, the inverse of the transpose, which is equal to the transpose of the inverse, and hence the use of the common shorthand where the order does not matter.

10. The motivation behind this choice is biorthogonality, roughly, a strong form of linear independence. In one dimension ($\mathbb{R}^1$), the only way for two points to be orthogonal is for one of them to be zero, by usual multiplication. In general, orthgonality takes a more relative face: two points (in $\mathbb{R}^n$) can be orthogonal relative to each other, without the constraint that at least one of them is orthogonal to all other points. Biorthogonality of a system of two sets of points simply says that each point from one set is orthogonal to all except one from the other set. This is a minimal and natural way to characterize a mapping that guarantees preservation of linear independence.

With basis points relating to matrix columns, and given the rules of matrix multiplication, expressing this requires taking the transpose of one of the matrices and writing:

$$XX^T = I.$$

This is reflected in F which makes it so that a matrix is mapped to its biorthogonal.

11. Let us pickup our original change of basis example, but parametrizing through F. This means that we take out uncanonical representation in equation 1 and choose ϕ to be represented by our specified F. We then have

$$a = F(B)\underline{a'}_i = B^{-T}\underline{a'}_i = F(C)\underline{\alpha'}_i = (BL)^{-T}\underline{\alpha'}_i.$$

12. In contrast to equations 2 we now have:

$$\begin{cases} \underline{\alpha'_i} & = L^T\underline{a'_i} \\ C & = BL \end{cases}. \tag{3}$$

In this situation, the point is said to change covariantly, like the basis.

13. It is known that variance is invisible the concerned change-of-basis matrices are orthogonal, this is clear since in this case, $X^{-1} = X^T$ and the two cases become indistinguishable.

14. Can we obtain point covariance in a simpler way, foregoing the example of dual basis? The obvious idea that comes to mind is to use

$$F(X) = X^{-1}.$$

This does not work out as nicely, and one ends up with

$$\underline{\alpha'_i} = BLB^{-1}\underline{a'_i}.$$

This beautifully illustrates another 'reason' for the transposition in the biorthogonality matrix and motivates the proper and common abstract approach, which lifts us from the flattened state of affairs of working in matrices instead of linear maps. Nevertheless, it is useful to have seen a more 'computational' reason for it.

15. More ambitiously, can we go the other way round and, requiring that we obtain the simplest covariant relation

$$\underline{\alpha'_i} = L\underline{a'_i},$$

and working out what that imposes on F? Here, we meet another beautiful motivation for mutlilinear algebra proper. We used a transpose, but is that a 'linear' thing? Is that a linear map proper? Even a short look tells us that something is wrong.

If we say that a transpose is a linear map proper, we should be able to to find a matrix T such that

$$X^T = TX.$$

If one examines what matrix multiplication constrains us with, one finds that there is no such T in general. Nevertheless, transposition can be seen as a linear map proper if one flattens the transposed matrix $(n \times n)$ into a $(n^2 \times 1)$ matrix. But then, this 'flattened' matrix cannot be used to multiply the coordinates. This again motivates linear and multilinear algebra proper. In tensor notation, the transpose can be written as

$$M^T = \sum_{i=1}^{n}(e_i^T \otimes I_n)S_{n,n}(I_n \otimes M)\left(\sum_{i=1}^{n}(e_j \otimes e_j)\right)e_i^T.$$ [2]

[2] http://math.stackexchange.com/questions/1143614, http://math.stackexchange.com/questions/125862

III Appendix

III.1 Nonlinear Parametrization

In our uncanonical vector representation, we constrained the map ϕ to be (multi)linear. What is wrong with a nonlinear continuous bijective one? This is almost never treated because it is not considered a useful avenue. Here is a simple reason for that. Let us take a one-dimensional vector space V with a vector a, a basis $B = \{b\}$ and another $B' = \{b'\}$. Let us also take a non-linear function $f : V \to V$ acting in the following way: $f(x) = (\text{Rep}_B(x))^3 b$, clearly, f is bijective so at least when applied to all vectors in V one gets another sets that spans V. For contrast, Let l be some linear bijection, $l : V \to V$. Let $y = l(a)$. Now y is related to a, and this relation transfers nicely to their coordinates: $y = l(a) = l(a_1 b) = a_1 l(b)$. If $b' = l(b)$, then we can use $y = a_1 b'$. The non-usefulness comes from the fact that the coordinates 'break', become 'useless', when we consider the non-linear function. $z = f(a) = f(a_1 b) \neq a_1 f(b)$. If $b' = f(b)$, it gives us nothing we can ***use***, since $z \neq a_1 b'$. In short a non-linear change of basis takes us out of linear algebra and the crucial identification between vector and coordinates becomes unworkable in the linear algebraic context.