THE LINEAR ALGEBRA PRIMER MULTILINEAR REGRESSION

Maurice J. Dupré
Department of Mathematics
New Orleans, LA 70118
email: mdupre@tulane.edu
27 April 2009

Abstract. We give an elementary treatment of linear algebra and apply it to give an elementary devlopment of multilinear regression.

1. INTRODUCTION

Mathematics is basically the art of inventing ways to handle information. For instance it is often said that a picture is worth a thousand words, and in mathematics the right picture can often be the key to solution of a problem. When it comes to handling numerical information, the right pictures usually come from linear algebra is it seems to provide an optimal blend of pictorial and numerical information in a way that is extremely convienient for transitioning back and forth between the numerical and the pictorial. As well, the most applicable area of modern mathematics certainly has to be linear algebra. In almost every area of mathematics, when the theory is developed to the point that actual computations can be done, we find linear algebra at the point where the rubber meets the road. Consequently, a good basic understanding of linear algebra should be part of every scientist's arsenal. Here, we will give an elementary development of the basic ideas and use them to develop the theory of multilinear regression as applied in statistics. We will be very brief, so we will skip over some of the fundamentals which a proper treatment would include in order to consentrate on the facts needed for multilinear regression.

2. VECTOR SPACES AND LINEAR TRANSFORMATIONS

A vector space is basically a set V of objects called vectors which can be added and multiplied by real numbers, so that all the basic rules for dealing with vectors as arrows in space hold (see the document titled VECTORS on my website for a pictorial introduction to vectors as arrows in space). Thus, vector addition is associative and commutative, there is a zero vector, 0_V, which added to any vector cannot change that vector, and each vector v has a negative, denoted $-v$, so a vector added to its own negative results in the zero vector. In symbols, we have

$$(u+v)+w = u+(v+w), \text{ for any vectors } u, v, w \text{ belonging to } V,$$

$$u+v = v+u, \text{ for any vectors } u, v \text{ belonging to } V,$$

$$v + 0_V = v, \text{ for any vector } v \text{ belonging to } V,$$

$$v + (-v) = 0_V, \text{ for any vector } v \text{ belonging to } V.$$

Moreover, regarding the multiplication of vectors by real numbers, we assume that it is associative and distributive with respect to vector addition and that multiplication by the number 1 will not change a vector. In symbols we have

$$r(sv) = (rs)v, \text{ for any real numbers } r, s \text{ and any vector } v \text{ belonging to } V,$$

$$(r + s)v = rv + sv, \text{ for any real numbers } r, s \text{ and any vector } v \text{ belonging to } V,$$

$$r(v + w) = rv + rw, \text{ for any real number } r \text{ and any vectors } v, w \text{ belonging to } V,$$

$$1v = v, \text{ for any vector } v \text{ belonging to } V.$$

With these rules, all the usual things we do with vectors as arrows in space will work. For instance, we can see

$$\text{if } v \text{ belongs to } V, \text{ and if } v + v = v, \text{ then } v = 0_V,$$

by simply adding $-v$ to both sides of the equation $v + v = v$. Then we see that for any vector v the result of multiplying it by the number zero is 0_V, since $0v + 0v = (0 + 0)v = 0v$, so

$$0v = 0_V, \text{ for any vector } v \text{ belonging to } V.$$

We call a real number a scalar when dealing with vectors and vector spaces, so the multiplication of a vector by a real number is usually called **scalar multiplication**.

If $U \subset V$, where V is a vector space, and if for any vectors v, w belonging to U and any real number r it is true that $v + rw$ again belongs to U, then U is called a **vector subspace** or **linear subspace** of V. Notice that by taking $r = 1$, we have $v + w$ belongs to U whenever v and w belong to U, and taking $v = 0$, we see that rw belongs to U whenever r is any real number and w belongs to U. Conversely, if these two cases apply to U, that is, if the vector sum of any two vectors in U is again in U, and if any scalar multiple of a vector in U is again in U, then U is a subspace of V. Now it is easy to see that U is a vector space in its own right as all the rules will apply since they apply in $V \supset U$.

For the vectors as arrows in space there is also the **dot product** of two vectors which results in a number. A **Euclidean Space** is a vector space E in which we can form the dot product of any two vectors v, w which is usually denoted $v \cdot w$. Here again, we assume the usual rules that work for vectors as arrows in space:

$$u \cdot (v + w) = u \cdot v + u \cdot w, \text{ for any vectors } u, v, w \text{ belonging to } E,$$

$$v \cdot w = w \cdot v, \text{ for any vectors } v, w \text{ belonging to } E,$$

$$(rv) \cdot w = r(v \cdot w), \text{ for any real number } r \text{ and any vectors } v, w \text{ belonging to } E.$$

$$v \cdot v \geq 0, \text{ for any vector } v \text{ belonging to } E,$$

$$\text{for any vector } v \text{ in } E, \text{ if } v \cdot v = 0, \text{ then } v = 0_E.$$

We define the **length** of the vector v in E denoted $\|v\|$, by setting

$$\|v\| = \sqrt{v \cdot v}, \text{ for any vector } v \text{ belonging to } E.$$

It can be shown using a little algebra and the rules here that if v, w belong to the Euclidean space E, then the **Cauchy-Schwarz** Inequality holds:

$$|v \cdot w| \leq \|v\| \cdot \|w\|.$$

If we think of vectors as arrows in space, for two such arrows having their tails at a common point, their heads are connected by the difference vector whose length is then the distance between the two vectors heads. Thus, if v and w are vectors, then thinking in terms of head to tail addition, $w - v$ is a vector which if situated so its tail is at the head of v, then its head coincides with the head of w. We therefore see that the distance from the head of v to the head of w is $\|w - v\|$. Therefore we have a way to compute distances in any Euclidean space, so Euclidean spaces are geometric things. If $v \cdot w = 0$, we say that v and w are perpendicular, as this is certainly the case for arrows in space. In general, these rules imply that if v and w belong to the Euclidean space E, then

$$\|v + w\|^2 = (v + w) \cdot (v + w) = (v \cdot v) + (w \cdot w) + 2(v \cdot w) = \|v\|^2 + \|w\|^2 + 2(v \cdot w),$$

so in particular,

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2, \text{ if } v \text{ and } w \text{ are perpendicular},$$

which of course is the **Pythagorean Theorem**.

Another useful fact about Euclidean spaces, is that often equations involving vectors can be reduced to numerical equations via the following trick. First notice that if v is any vector in the Euclidean space E, and if $v \cdot w = 0$ for every vector w in E, then we can take $w = v$ and conclude that $v \cdot v = 0$. But, one of the basic assumptions says that if $v \cdot v = 0$, then $v = 0$. Now suppose that we have a vector equation $u = v$ which may or may not be true. In order to see that it is true, we just take the dot product of both sides of the equation with an arbitrary vector w in E and see if a true equation results. If so, the equation $u = v$ must be true. Indeed, if $u \cdot w = v \cdot w$ for every vector w in E, then

$$(u - v) \cdot w = u \cdot w - v \cdot w = 0, \text{ for every vector } w \text{ in } E,$$

and therefore $u - v = 0$, so $u = v$.

If V and W are any vector spaces, a **linear transformation** $T : V \longrightarrow W$ is a function from V to W such that

$$T(v + rw) = T(v) + rT(w), \text{ for any real number } r \text{ and any vectors } v, w \text{ belonging to } V.$$

Taking $r = 1$ gives $T(v + w) = T(v) + T(w)$, so T is **additive** if T is linear. Also, taking $v = 0_V$ gives $T(rw) = rT(w)$, so T commutes with scalar multiplication, that is T is **homogeneous**. Clearly if T is any function from V to W which is both additve and homogeneous, then T is linear.

It is very useful to notice that the composition of linear transformations is linear, that is, if U, V, W are all vector spaces, if $S : U \longrightarrow V$ and $T : V \longrightarrow W$ are both linear transformations, then their composition $[T \circ S] : U \longrightarrow W$ is also a linear transformation. It is customary to denote the composition of linear transformations simply by juxtaposition,

$$T \circ S = TS,$$

so in the future we will simply write TS for the composition. If V is any vector space, then the **identity map** of V is denoted $Id_V : V \longrightarrow V$ and then $I_V(x) = x$, for any vector x in V. Obviously, if $T : V \longrightarrow W$ is any function, then

$$[Id_W]T = T = T[Id_V].$$

We say that $T : V \longrightarrow W$ and $S : W \longrightarrow V$ are **mutually inverse** functions if

$$S \circ T = Id_V \text{ and } T \circ S = Id_W.$$

In this case, S is completely determined by T, and T is completely determined by S, so we write

$$S = T^{-1} \text{ and } T = S^{-1},$$

and we say that T has an inverse or is invertible, and we call S the **inverse** of T. Notice here that

$$S(T(v)) = [S \circ T](v) = Id_V(v) = v \text{ and } T(S(w)) = [T \circ S](w) = Id_W(w) = w,$$

so each "undoes" what the other "does". If we have two invertible functions $S : U \longrightarrow V$ and $T : V \longrightarrow W$, then their composition is invertible and we have

$$[T \circ S]^{-1} = [S^{-1}] \circ [T^{-1}].$$

Of course this applies to linear transformations, so if S and T are linear, this would be written

$$[TS]^{-1} = S^{-1}T^{-1}.$$

If E and F are Euclidean spaces and $T : E \longrightarrow F$ is a linear transformation such that $\|T(v)\| = \|v\|$, for every vector v in E, then we call T a **linear isometry**. In this case, we see from linearity that T preserves all distances. If T is a linear isometry of E into F, then for any vectors v and w in E, we have

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2(v \cdot w)$$

and

$$\|T(v) + T(w)\|^2 = \|T(v)\|^2 + \|T(w)\|^2 + 2([T(v)] \cdot [T(w)])$$

and as all the corresponding terms involving the squared lengths are the same for both equations due to the fact that T is isometric, it follows that the dot product terms must also be the same which means

$$[T(v)] \cdot [T(w)] = v \cdot w, \text{ for any } v, w \text{ in } E \text{ if } T : E \longrightarrow F \text{ is isometric.}$$

3. EXAMPLES OF VECTOR SPACES AND LINEAR TRANSFORMATIONS

The simplest example of a vector space is the case where V is a one element set whose only member we denote by 0_V. In this case it is easy to see that all the rules apply in a very trivial way. We will sort of abuse the set theory notation a little bit here and just write $V = 0_V$, or even simply $V = 0$ for this example. We can even see that 0 is a Euclidean space, since the dot product of any two vectors here is just the real number zero. Notice that if V is any vector space, then 0_V is a vector subspace of V. Of course V itself is a vector subspace of V.

We denote by $\mathbb{R}$ the set of all real numbers, and we can observe that $\mathbb{R}$ is a vector space where we use the ordinary addition of numbers for the vector addition and ordinary multiplication of numbers for the scalar product. We also can define the dot product of two numbers to be the ordinary product, so with these definitions, $\mathbb{R}$ becomes a Euclidean space, where the scalar product and dot product coincide.

If V and W are vector spaces, we can make the cartesian product $V \times W$ into a vector space. Recall that the cartesian product of two sets $V \times W$ is the set of ordered pairs (v, w) where v belongs to V and w belongs to W. We can define the sum of ordered pairs by setting

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1, w_2), \text{ for any } v_1, v_2 \text{ in } V \text{ and for any } w_1, w_2 \text{ in } W.$$

Similarly we define the scalar multiplication in $V \times W$ by setting

$$r(v, w) = (rv, rw), \text{ for any } r \text{ in } \mathbb{R} \text{ and any } v \text{ in } V \text{ and any } w \text{ in } W.$$

We call the entries of the pair the coordinates here, so we say the operations here are simply defined coordinatewise. Obviously we could do the same thing with more coordinates or any number of coordinates. If $V_1, V_2, V_3, ..., V_n$ are all vector spaces, then we can form the cartesian product V of all of them, so

$$V = V_1 \times V_2 \times V_3 \times ... \times V_n$$

which consists of sequences $(v_1, v_2, v_3, ..., v_n)$ where v_k is required to belong to V_k for each $k \leq n$. Such sequences of length n are called $n-$tuples. Again, we define the vector space operations coordinatewise, so

$$(v_1, v_2, v_3, ..., v_n) + (w_1, w_2, w_3, ..., w_n) = (v_1 + w_1, v_2 + w_2, v_3 + w_3, ..., v_n + w_n)$$

and

$$r(v_1, v_2, v_3, ..., v_n) = (rv_1, rv_2, rv_3, ..., rv_n).$$

It is easy to see that the cartesian product $V = V_1 \times V_2 \times ... \times V_n$ is a vector space as the operations being defined corrdinatewise satisfy the rules in each coordinate.

If each vector space V_k with $k \leq n$ is actually a Euclidean space, then V, the cartesian product is also a Euclidean space, where we define the dot product of two $n-$tuples by setting

$$(v_1, v_2, v_3, ..., v_n) \cdot (w_1, w_2, w_3, ..., w_n) = [v_1 \cdot w_1] + [v_2 \cdot w_2] + [v_3 \cdot w_3] + ... + [v_n \cdot w_n].$$

In particular, if $V_k = W$ for each $k \leq n$, then we denote the cartesian product by W^n. Thus, $\mathbb{R}^n$ is a Euclidean space as $\mathbb{R}$ is a Euclidean space, and W^n is a Euclidean space if W is.

If V and W are vector spaces and $T : V \longrightarrow W$ is a function, then for $U \subset V$, we can define $T(U)$ to be the set of vectors of the form $T(u)$ where u comes from the set U. Then $T(U)$ is a subset of W. If U is a vector subspace of V and T is linear, then we can observe that $T(U)$ is a vector subspace of W. To see this, notice that if w_1, w_2 belong to $T(U)$, then there are vectors u_1 and u_2 belonging to U so that $w_1 = T(u_1)$ and $w_2 = T(u_2)$. It follows from the linearity of T that for r in $\mathbb{R}$ we have, setting $u = u_1 + ru_2$, that

$$w_1 + rw_2 = T(u_1) + rT(u_2) = T(u_1 + ru_2) = T(u),$$

and as U is now assumed to be a vector subspace of V, it follows that u belongs to U and therefore $w_1 + rw_2$ belongs to $T(U)$, showing that $T(U)$ is indeed a vector subspace of W.

As a useful example of this, if V is any vector space and $v = (v_1, v_2, v_3, ..., v_n)$ belongs to V^n, then we can define a linear transformation denoted $T_v : \mathbb{R}^n \longrightarrow V$, by

$$T_v(r_1, r_2, r_3, ..., r_n) = r_1v_1 + r_2v_2 + r_3v_3 + ... + r_nv_n.$$

It is not hard to see that T_v is actually a linear transformation. In this case, we can take the subspace $U = \mathbb{R}^n$ and see that $T_v(\mathbb{R}^n)$ is a vector subspace of V which we call the **linear span** of the vectors in the $n-$tuple $v = (v_1, v_2, v_3, ..., v_n)$. We say that V is a **finite dimensional vector space** if V is the linear span of a finite tuple of vectors. The smallest n for which there is an $n-$tuple v in V^n with $T_v(\mathbb{R}^n) = V$ is called the **Dimension** of V and denoted by $dim(V)$. The $n-$tuple is then called a **frame** for V.

Suppose now that $T : V \longrightarrow W$ is any function and $U \subset W$. We define the set $T^{-1}(U)$ as the set of all members v of V for which $T(v)$ belongs to U. If V and W are vector spaces and T is linear, and if U is a vector subspace of W, then $T^{-1}(U)$ is a vector subspace of V. To see this, if v and w belong to $T^{-1}(U)$ and r is a real number, then $T(v)$ and $T(w)$ belong to U and therefore so does $T(v) + rT(w) = T(v + rw)$ which means that $v + rw$ is in $T^{-1}(U)$. In particular, since 0_W is a vector subspace of W, the set $T^{-1}(0_W)$ is a vector subspace of V called the **kernel** or the **null space** of T.

4. THE ADJOINT OPERATION ON LINEAR TRANSFORMATIONS

If E is a finite dimensional Euclidean space, then it can be shown that there is a frame $u = (u_1, u_2, u_3, ..., u_n)$ where $n = dim(E)$ such that $u_i \cdot u_j = 1$ or 0 according to whether or not $i = j$. Thus, each vector in the frame has unit length and they are all mutually perpendicular to each other. You have to admit that calling n the dimension of E seems forced here. In this case, if we form the transformation $T_v : \mathbb{R}^n \longrightarrow V$, then we can show that T_v is an isometry. To see this, notice that if e_k is the vector in $\mathbb{R}^n$ whose k^{th} coordinate is 1 and all of whose other coordinates are zero, then $e = (e_1, e_2, e_3, ..., e_n)$ is an orthonormal frame for $\mathbb{R}^n$. Moreover, $T_u(e_k) = u_k$ for each $k \leq n$. If v is any vector in E, then since $T_u(\mathbb{R}^n) = E$, it follows that there is an $n-$tuple of real numbers $(r_1, r_2, r_3, ..., r_n)$ in $\mathbb{R}^n$ such that $T_u(r_1, r_2, r_3, ..., r_n) = v$. This means that

$$v = r_1u_1 + r_2u_2 + r_3u_3 + ... + r_nu_n.$$

To see how to compute these real number coefficients with the dot product, first take the dot product of both sides of the equation with u_1 and notice that as u_1 is perpendicular to $u_2, u_3, ..., u_n$ then all the terms of the right hand side dot product vanish except the first and we have

$$v \cdot u_1 = r_1(u_1 \cdot u_1) = r_1 1 = r_1.$$

If we dot both sides of the equation for v with u_2 we see that as u_2 is perpendicular to u_1 and perpendicular to $u_3, u_4, ..., u_n$, all terms except the second term vanish and

$$v \cdot u_2 = r_2(u_2 \cdot u_2) = r_2 1 = r_2.$$

Likewise we see that

$$v \cdot u_k = r_k(u_k \cdot u_k) = r_k 1 = r_k, \text{ for any } k \leq n.$$

That is finally,

$$v \cdot u_k = r_k, \text{ for any } k \leq n.$$

In other words, we can say that

$$v = (v \cdot u_1)u_1 + (v \cdot u_2)u_2 + (v \cdot u_3)u_3 + ... + (v \cdot u_n)u_n, \text{ for every vector } v \text{ in } E.$$

Therefore, if w is another vector in E, then we likewise have

$$w = (w \cdot u_1)u_1 + (w \cdot u_2)u_2 + (w \cdot u_3)u_3 + ... + (w \cdot u_n)u_n.$$

If we form $v \cdot w$, we see that

$$v \cdot w = [r_1 u_1 + r_2 u_2 + r_3 u_3 + ... + r_n u_n] \cdot w = r_1(w \cdot u_1) + r_2(w \cdot u_2) + r_3(w \cdot u_3) + ... + r_n(w \cdot u_n)$$

so

$$v \cdot w = (v \cdot u_1)(u_1 \cdot w) + (v \cdot u_2)(u_2 \cdot w) + (v \cdot u_3)(u_3 \cdot w) + ... + (v \cdot u_n)(u_n \cdot w).$$

in particular,

$$\|v\|^2 = v \cdot v = r_1^2 + r_2^2 + r_3^2 + ... + r_n^2 = \|(r_1, r_2, r_3, ..., r_n)\|^2.$$

which means T_u is isometric, as $v = T_u(r_1, r_2, r_3, ..., r_n)$. As a consequence, we have

$$[T_u(r)] \cdot [T_u(s)] = r \cdot s, \text{ for any } r, s \text{ in } \mathbb{R}^n.$$

Suppose that E is a finite dimensional Euclidean space with orthonormal frame $u = (u_1, ..., u_n)$. Suppose that $S : E \longrightarrow \mathbb{R}$ is a linear map. Then $S(u_k)$ is a number for each $k \leq n$, so that the $n-$tuple $(S(u_1), S(u_2), S(u_3), ..., S(u_n))$ is a vector v belonging to $\mathbb{R}^n$. Thus, we define

$$v_S = (S(u_1), S(u_2), S(u_3), ..., S(u_n)).$$

Then for any vector w in E we have

$$w = (w \cdot u_1)u_1 + (w \cdot u_2)u_2 + (w \cdot u_3)u_3 + ... + (w \cdot u_n)u_n,$$

and therefore

$$S(w) = (w \cdot u_1)S(u_1) + (w \cdot u_2)S(u_2) + (w \cdot u_3)S(u_3) + ... + (w \cdot u_n)S(u_n)$$

$$= w \cdot [S(u_1)u_1 + S(u_2)u_2 + S(u_3)u_3 + ... + S(u_n)u_n] = w \cdot [T_u(v_S)].$$

This means that

$$S(w) = w \cdot [T_u(v_S)], \text{ for every vector } w \text{ in } E.$$

Finally, set $u_S = T_u(v_S)$. We then have simply

$$S(w) = w \cdot u_S, \text{ for every vector } w \text{ in } E.$$

This is a special case of the **Riesz Representation Theorem** which states in particular that if E is a finite dimensional Euclidean space and if $S : E \longrightarrow \mathbb{R}$ is a linear transformation, then there is a unique vector u_S in E such that

$$S(v) = v \cdot u_S, \text{ for every vector } v \text{ in } E.$$

Of course the uniqueness follows from the fact that if w_S is also a vector in E with $S(v) = v \cdot w_S$ for every v in E, then $u_S \cdot v = w_S \cdot v$ for every vector v in E, and therefore $u_S = w_S$.

Suppose now that $T : E \longrightarrow F$ is any linear transformation of Euclidean spaces. We say that $U : F \longrightarrow E$ is **adjoint** to T provided that

$$[T(v)] \cdot w = v \cdot [U(w)], \text{ for every } v \text{ in } E \text{ and every } w \text{ in } F.$$

If U_1 and U_2 are both adjoint to T, and w is a vector in F, then for every vector v in E, we have $v \cdot [U_1(w)] = T(v) \cdot w = v \cdot [U_2(w)]$, so by our usual trick we see that $U_1(w) = U_2(w)$. It follows that $U_1 = U_2$. If w belongs to F, then to find $U(w)$, we can notice that we can define a linear transformation $S_w : E \longrightarrow \mathbb{R}$ by the rule $S_w(v) = [T(v)] \cdot w$, for every v in E. Assuming a case of the Riesz Representation Theorem applies, there is a unique u in E such that $S_w(v) = v \cdot u$ for each v in E. If U is adjoint to T, then

$$v \cdot [U(w)] = [T(v)] \cdot w = S_w(v) = v \cdot u, \text{ for any } v \text{ in } E,$$

and therefore $U(w) = u$. This means that the Riesz Representation Theorem guarantees the existence of a function adjoint to T, and we merely need to check that U defined by this process is linear. We will leave the details of this to the interested reader. We see in particular that T must have an adjoint if E is finite dimensional, even if F is not, by our special case of the Riesz Representation Theorem.

If $T : E \longrightarrow F$, we see that there is at most one adjoint for T and if E is finite dimensional, then T has an adjoint. We denote this depence on T by writing T^* for the adjoint of T, when it exists, so

$$[T(v)] \cdot w = v \cdot [T^*(w)], \text{ for every } v \text{ in } E \text{ and every } w \text{ in } F.$$

Clearly, if T has an adjoint then so does T^* and it is T itself, that is

$$[T^*]^* = T.$$

If $T : E \longrightarrow F$ and $U : F \longrightarrow G$ both have adjoints, then so does UT and

$$[UT]^* = T^* U^*.$$

To see this, we note that if v is in E and if w is in G, then

$$[UT](v) \cdot w = [U(T(v))] \cdot w = T(v) \cdot U^*(w) = v \cdot [T^*(U^*(w))] = v \cdot ([T^*U^*](w)),$$

showing that if T and U both have adjoints, then so does UT and indeed, by the uniqueness of adjoints we have $[UT]^* = T^*U^*$ as claimed.

5. PROJECTION OPERATORS

Suppose that B is any set and $R : B \longrightarrow B$ is any function. We say that R is a **retraction mapping** or simply a **retraction** if $R \circ R = R$. We say that b in B is a **fixed point** of R if $R(b) = b$. Suppose that R is a retraction and let F be the set of all values of R, so $F = R(B)$. This means that v belongs to F if and only if there is some b in B such that $v = R(b)$. Clearly any fixed point of R must be a value of R and so F contains the set of all fixed points of R. But, for R a retraction we have that v in F means there is b in B such that $v = R(b)$, so using $R \circ R = R$, we have $R(v) = R(R(b)) = [R \circ R](b) = R(b) = v$, showing that v is a fixed point. Thus, $F = R(B)$ is the set of all fixed points of R.

Suppose now that V is a vector space and that $R : V \longrightarrow V$ is a linear retraction. Thus R is simultaneously a retraction and a linear transformation. Then $W = R(V)$ is a vector subspace of V and $K = R^{-1}(0_V)$ is a vector subspace of V. For the moment let's fix attention on a vector v in V. Set $w = R(v)$. Then we know that w is a fixed point of R, so $R(w) = w$. Define $u = v - w$. Then $R(u) = R(v-w) = R(v) - R(w) = w - w = 0_V$, so u belongs to K. Thus, we have

$$v = u + w, \text{ with } u \text{ in } K \text{ and } w \text{ in } W.$$

That is, we have every vector in V is the sum of two vectors, one from K and one from W. We denote this by simply writing

$$V = K + W.$$

On the other hand, if v belongs to both K and W, then we have both $R(v) = 0_V$ and $R(v) = v$ from which we conclude that $v = 0_V$, so we denote this by simply writing

$$K \cap W = 0_V.$$

In this situation, we say that we have **split** the vector space V as the direct sum of the subspaces K and W. We say that R is the **projection of** V **onto** W **along** K.

Suppose now that $R : E \longrightarrow E$ is a linear retraction of the Euclidean space E. Further, suppose that R is self-adjoint which is to say it is its own adjoint, so

$$[R(v)] \cdot w = v \cdot [R(w)], \text{ for any } v, w \text{ in } E.$$

As before, define $K = R^{-1}(0_E)$ and $W = R(E)$. In this case, if u is in K and w is in W, then $R(u) = 0_E$ and $R(w) = w$, so

$$u \cdot w = u \cdot [R(w)] = [R(u)] \cdot w = 0_E \cdot w = 0.$$

From this we conclude that every vector in K is perpendicular to every vector in W which means that K and W are perpendicular subspaces, but as before, $K + W = E$. We have therefore split E as the direct sum of two orthogonal subspaces.

An important general problem in mathematics is the problem of finding the point of a subset closest to a given point not in the subset. If we are in ordinary space, we know that to find the closest point on a plane to a given point not on the plane, we drop a perpendicular to the plane. That is, we project to the plane along the line perpendicular to the plane. Thus, if R is a self-adjoint projection, then for $W = R(E)$, if v is any vector in E, then $R(v)$ should be the closest point of W to v. To see this, if z is any point of W, then $R(z) = z$ and we can write $v = u + w$ with u in K and w in W, so $R(v) = w$, and $v - R(v) = v - w = u$, which is perpendicular to W since it is in K. Consider now the vector $v - z$ whose length is the separation distance between z and v. We have $v - z = (u + w) - z = u + (w - z)$. But, both w and z belong to the subspace W and therefore $w - z$ belongs to W. On the other hand, u belongs to K which is perpendicular to W. Thus, by the Pythagorean Theorem in Euclidean space,

$$\|v - z\|^2 = \|u + (w - z)\|^2 = \|u\|^2 + \|w - z\|^2 \geq \|u\|^2,$$

and we see that no matter how z in W is chosen, the distance to v is at least $\|u\|$ which is the length of the vector $u = v - w = v - R(v)$, and therefore $R(v)$ is the closest point of W to v. For if we make the choice $z = R(v)$, then $w = z$ so $w - z = 0_E$ and $\|v - z\| = \|u\|$.

Suppose that E is a Euclidean space and F is a subspace of E, which is finite dimensional. Suppose that $u = (u_1, u_2, u_3, ..., u_n)$ is an orthonormal frame for F. We can define the linear map $R : E \longrightarrow E$ by

$$R(v) = (v \cdot u_1)u_1 + (v \cdot u_2)u_2 + (v \cdot u_3)u_3 + ... + (v \cdot u_n)u_n.$$

It is not hard to see that R is linear. We know that since u is an orthonormal frame for F, that if v is in F, then $R(v) = v$. On the other hand, it is clear that for any v in E the the vector $R(v)$ belongs to the subspace

F. Thus, if v is in E, then $R(R(v)) = R(v)$ which means $RR = R$ and R is a linear retraction. If v and w belong to E, then

$$[R(v)] \cdot w = (v \cdot u_1)(w \cdot u_1) + (v \cdot u_2)(w \cdot u_2) + ... + (v \cdot u_n)(w \cdot u_n) = v \cdot [R(w)].$$

6. LINEAR REGRESSION

Suppose that E is a finite dimensional Euclidean space, that A is any Euclidean space, and that $T : E \longrightarrow A$ is a linear map. What we have in mind is that A is an algebra of unknowns in some setting or with more restrictions, an algebra of random variables, and in typical applications, $T = T_v$ where v is an $n-$tuple of vectors in A. Given v in A what we wish to do is find b in E so that $T(b)$ is the closest point of $T(E) \subset A$ to v and therefore we would regard this closest point as the best approximation of v if we are restricted to using points in $T(E)$ to approximate v. Of course, if $R : A \longrightarrow A$ is a self-adjoint linear retraction with $R(A) = T(E)$, then we just take $R(v)$ since it is the closest point of $T(E)$ to v. We then would need to solve $R(v) = T(x)$ and the solution for x would give the point b of E for which $T(b)$ is closest to v. In order to define R, as we have seen, we could find an orthonormal frame for $T(E)$. To do this, we take any frame for E and look at the values in $T(E)$ and reduce to a frame and then construct an orthonormal frame. This is extremely tedious and inefficient. Here is where our theory comes to the rescue. First, we keep in mind that if b is the solution, then $R(T(b)) = T(b)$ and $v - T(b)$ is perpendicular to $T(E)$. Thus, for any vector e in E, we have

$$0 = [v - T(b)] \cdot [T(e)] = [T^*(v - T(b))] \cdot e = [T^*(v) - T^*T(b)] \cdot e,$$

and because this equation holds for every e in E, it follows that

$$T^*(v) - T^*T(b) = 0_E$$

which means that

$$T^*(v) = T^*T(b).$$

At this point, we need to assume that T^*T is invertible, so then we have

$$b = [T^*T]^{-1}T^*(v).$$

Notice this means that $R(v) = T(b) = T[T^*T]^{-1}T^*(v)$, so

$$R = T[T^*T]^{-1}T^*, \text{ assuming } T^*T \text{ invertible }.$$

We can see directly that $RR = R$, since

$$RR = T[T^*T]^{-1}T^*T[T^*T]^{-1}T^* = T([T^*T]^{-1}[T^*T])[T^*T]^{-1}T^* = T[T^*T]^{-1}T^* = R.$$

Also, it is easy to see that R is self adjoint. For instance, T^*T is obviously self-adjoint, and if S is any invertible linear transformation which has an adjoint, then S^{-1} has an adjoint and

$$[S^{-1}]^* = [S^*]^{-1}.$$

Now this means that $[T^*T]^{-1}$ is self-adjoint, and if S is any self-adjoint operator, then

$$[TST^*]^* = [T^*]^*ST^* = TST^*,$$

and thus in particular, R is self-adjoint.

DEPARTMENT OF MATHEMATICS, TULANE UNIVERSTIY, NEW ORLEANS, LA 70118
E-mail address: mdupre@tulane.edu