

Linear Algebra

Nathaniel Beck
Department of Political Science
University of California, San Diego
La Jolla, CA 92093
beck@ucsd.edu
<http://weber.ucsd.edu/~nbeck>

April, 2001

Vectors

k-vector as point in Euclidean k dimensional space (R^k)

$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is a point in R^2 .

Note that for reasons to clear later, we always represent vectors as COLUMN vectors

Define *addition* of two vectors as the sum element by element, so

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \end{pmatrix} \quad (1)$$

where addition is only defined for two vectors in same dimensional space (so both must be k-vectors)

Define *scalar multiplication* as the element by element product of the scalar with each element of the vector

$$a \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2a \\ 3a \end{pmatrix} \quad (2)$$

Linear combination of vectors x and y is $ax + by$ for any scalars a and b .

Consider a set of vectors. The span of this set is all the linear combinations of the vectors.

Thus the span of $\{x, y, z\}$ is the set of all vectors of the form $ax + by + cz$ for all scalars (real numbers) a, b, c .

Basis, dimension

Consider a set of vectors, X . Suppose X is the span of some vectors a_1, \dots, a_n . The latter set is a *basis* for X .

There are lots of bases for X . Often one has really nice features and so we use it.

Particularly interested in basis of R^n . Consider R^2 . A nice basis is the two vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Note that any vector in R^2 , say $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ so we have a basis.

But lot of others, say $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Need to show that

$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ for some scalars c_1 and c_2 .

To see we can do this, just solve the two linear equations. Note that vectors and solving linear equations intimately related.

Note that if $\{x, y\}$ is a basis, so is $\{x, y, z\}$ for any z since we can just write $a = a_1x + a_2y + 0z$ since $\{x, y\}$ is a basis.

A set of vectors is *linearly dependent* if there is some *non-trivial* linear combination of the vectors that is zero. By non-trivial we mean a combination other than with all scalars being zero, since $0x + 0y = 0$ of course.

A set of vectors is *linearly independent* if it is not linearly dependent.

Note that if a set of vectors is linearly dependent, we have, say, $ax + by = 0$ for non-zero a or b , so we can write $y = \frac{a}{b}x$, that is, one vector is a linear combination of other vectors in the set.

Since we can always rewrite a vector that is a linear combination of a linearly dependent set of vectors as a new linear combination of the linearly independent subset of the vectors. Thus we can say a nice basis consists only of linearly independent vectors.

We also particularly like the "natural" basis of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ where each vector is of unit length and the two are orthogonal (at right angles). But, as noted, tons of other bases, even if restrict to linear dependence. Let us call the vectors in the natural basis i_1, i_2 and similarly for higher dimensional spaces.

Note that two vectors form a basis for R^2 , and 2 is the dimension of R^2 . This is not an accident, the dimension of a vector space is the number of linearly independent vectors needed to span it.

Note also that in R^2 that any three vectors must be linearly dependent, since we can write the third vector as a linear combination of the other two.

Again, just solve the equations to make the third vectors sum to zero non-trivially.

Note that any set of vectors that include the zero vector $\mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ must be linearly dependent since, say, $0x + a\mathbf{0} = \mathbf{0}$ for $a \neq 0$.

Matrices

A matrix is a rectangular array of numbers or a series of vectors placed next to each other. An example is

$$\begin{pmatrix} 0 & 1 & 3 & 6 \\ 2 & 7 & 3 & 4 \\ 2 & 1 & 2 & 0 \end{pmatrix} \quad (3)$$

Matrices are called $m \times n$ with m rows and n columns.

We can define scalar multiplication in the same way as for vector, just multiply each element of the matrix by the scalar.

We can define addition of matrices as element by element addition, though we can only add matrices of the same number and rows and column.

Can understand matrices by the following circuitous route:

A linear transform is a special type of transformation of vectors; these transforms map n dimensional vectors into m dimensional vectors, where the dimensions might or might not be equal.

A *linear transform*, T has the properties:

$$T(ax) = aT(x) \text{ where } a \text{ is a scalar} \quad (4)$$

$$T(x + y) = T(x) + T(y) \quad (5)$$

Note that if we know what T does to the basis vector, we completely know T since $T(x) = x_1T(i_1) + \dots + x_nT(i_n)$.

Let us then think of an $m \times n$ matrix as defining a linear transform from R^n to R^m where the first column of the matrix tells what happens to the i_1 , the second column to i_2 and so forth.

Thus the above example matrix defines a transform that takes i_1 to the vector $\begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$ and so forth.

Thus we can define Mx , the product of a matrix and a vector as the vector that the transform defined by M maps x into.

Say $M = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 0 & 1 \end{pmatrix}$ which is a linear transform from $R^2 \rightarrow R^3$. What happens to, say $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$? This vector is $1i_1 + 2i_2$ and so

$$T(x) = Mx = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \\ 2 \end{pmatrix} \quad (6)$$

This is the same thing we do woodenheadly by taking each element by taking the sum of the product of the elements of row i with the vector.

We could compute MN by thinking of N as a k column vectors, and the above operation would produce k column vectors with the same number of rows as M . Note that the number of columns in M (the dimension of the space we are mapping) from must equal the number of elements in each of the column vectors of N .

But we can also think of the matrix product MN as the composition of the transforms corresponding to M and N . Say M is $n \times m$ and N is $m \times k$.

So N takes k vectors into m vectors and M takes m into n vectors so we can see MN takes k vectors into n vectors, by first taking them into m vectors and then taking those m vectors into n vectors.

To understand the resulting product in matrix form, just think of it as telling us what happens to the k basis vectors after the two transformations. It is easy to see that the i, j th element of the product is given to us by taking the sum of the element by element products of row i and column j , and that we can only multiply matrices where the number of columns in the first one equals the number of rows in the second.

Note that matrix multiplication is not commutative, in that NM will usually not even be defined, since M takes m vectors into n vectors but N takes k vectors into m vectors, so N cannot operate on the vectors produced by M .

We can also think of $M + N$ as corresponding to the transform which has $M + N(x) = M(x) + N(x)$, always assuming the matrices are of the right dimension (in the jargon, *conformable*).

Square Matrices

First, define $\mathbf{0}_n$ as the $n \times n$ matrix containing all zeros and I_n as the square matrix ($n \times n$) with ones on the diagonal and zeros elsewhere. I_n is the *identity* matrix, and the subscript for dimension is often suppressed.

Note that $A + 0 = A$ (assuming all are $n \times n$ and that $A0 = 0$ so 0 behaves like a zero should).

Note also that $AI = IA = A$.

Thus, in terms of transforms, I corresponds to the linear transform from n space to itself which just leaves each vector alone, that is, $L(x) = x$. 0 corresponds to the transform which maps all n vectors into an n vector of all zeros.

Finally, if $x = 0$ (a column vector of all zeros), then we see that $Mx = 0$.

Are there other, non-zero, vectors that get mapped into zero? This turns out to be critical. Note that if $Mx = 0$ for non-zero x , then the elements of x give a non-trivial linear combination of the rows of M which add to zero, so the rows of M are linearly dependent.

If no non-zero x gets mapped to zero, then the rows of M are linearly independent.

We define the maximal number of linearly independent rows of M as its *row rank* and similarly for the *column rank*. It is easy to prove that the row and column ranks of a matrix are the same, so we can just talk of the rank of a matrix. If it is $n \times n$ its rank can be no bigger than n ; if it is n , then we say it has full rank, otherwise it is less than full rank.

Inverses of square matrices

Define the *inverse* of A , A^{-1} such that $AA^{-1} = A^{-1}A = I$. Inverse pop up all the time; question is how to find them and do they exist.

The computation is tedious. But we can understand by looking at the transform corresponding to A and it's inverse.

(Get computation by just solving for A^{-1} in the matrix equation $AA^{-1} = I$, nothing difficult, just tedious and not very insightful.

What is the inverse of a transform? It undoes the transform, that is, if we do a transform and then its inverse, we get back to the original vector. Note that the product of the matrix and its inverse is the identity matrix.

A necessary condition for the inverse to exist is that it be *one to one* and *onto*.

One to one means that each vector in the range of the transform is the transform of only one vector, that is, we do not have $L(x) = L(y)$ unless $x = y$. We need this since transforms take a vector into only one vector, and since the inverse would be a transform, it too must

take a vector into only one vector. But if $L(x) = L(y)$, what is the inverse of $L(x)$; is it x or y ?

Onto means that every vector in the range space is mapped by something, so that there is an x to make $y = L(x)$ for every y . If the transform is not onto, then some vectors in the range space have no map from the domain, so how do we then invert?

Note that $L(0) = 0$ since $L(0) = L(x - x) = L(x) - L(x) = 0$.

If any other points in the domain get mapped into zero, then the transform is not one to one, and so inverse does not exist.

If nothing other than 0 gets mapped into 0, then it is easy to show that the transform is one to one.

(Suppose not. Then we would have $L(x) = L(z) = y$, $x \neq z$, and so $L(x - z) = 0$ but $x - z \neq 0$.)

Thus to see if inverses exist, we need merely look at what gets mapped into zero (the *kernel* of the transform) and see if only zero gets mapped into zero.

Thus consider the transform $T(x_1, x_2) = (x_1)$ which takes $R^2 \rightarrow R^1$. Clearly all elements $(0, x_2)$ get mapped to 0, so the transform is not one to one and hence inverse does not exist.

Consider mapping $R^2 \rightarrow R^2$ by $L \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix}$. Note that $L(0) = 0$. Can $L(x) = 0, x \neq 0$? If so, both $x_1 + x_2 = 0, x_1 - x_2 = 0$. Solve and you will see that this holds only for $x_1 = x_2 = 0$. (Again, note the tie between matrices, linear transforms and solving linear equations.)

First, only square matrices can have inverses. If M is not square, it maps $R^n \rightarrow R^m$. If $m > n$, then more elements the transform cannot be onto (only onto transform could be onto an n dimensional subspace. If $m < n$, this discussion holds for the inverse.

So only square matrices have inverses. All we have to do is to check that only 0 gets mapped to 0.

But as we have seen, if something else gets mapped to zero, the columns of the matrix are linearly dependent.

Thus a square matrix has an inverse if it is full rank.

Determinant

Determinants are usually treated as annoying things to compute. But the determinant of A , $(|A|)$, is simply the hypervolume of the hyperrectangle defined by either the rows or columns of A .

Note that if one row or column is a linear combination of the others, then the hyperrectangle is not of full dimension, that is, has hypervolume of zero.

Thus a matrix has an inverse if and only if its determinant is non-zero.

Quadratic forms

Letting x be a vector and A a conformable matrix, we often deal with *quadratic forms* $x'Ax$ which is a scalar (x' is $1 \times n$, so product is 1×1 , that is, a scalar).

A matrix is positive definite (PD) if all quadratic forms involving it are positive (other than the trivial $x = 0$). Similarly for negative definite; positive semi-definite has all quadratic forms involving A being non-negative and similarly for negative semi-definite.

Note that quadratic forms are the matrix analogue of scalar quadratic equations. To see this, let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then

$$x'Ax = x' \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{pmatrix} = x_1^2 + 5x_1x_2 + 4x_2^2 \quad (7)$$

Matrix calculus

We need matrix calculus to do minima and maxima; matrix calculus is taking derivatives with respect to a vector x ; this is defined by the taking the derivatives with respect to each element of x and then treating all the derivatives as a vector.

Thus this is purely a notation saving device and the proofs are trivial.

$$\frac{dAx}{dx} = A' \quad (8)$$

$$\frac{dx'Ax}{dx} = 2Ax \quad (9)$$