

Inner product. Given a real vector space V , an inner product is defined to be a bilinear map $F : V \times V \rightarrow \mathfrak{R}$ such that the following holds:

- For all $v_1, v_2 \in V$, we have $F(v_1, v_2) = F(v_2, v_1)$. (Commutativity.)
- For all $v_1, v_2, v_3 \in V$, we have $F(v_1, v_2 + v_3) = F(v_1, v_2) + F(v_1, v_3)$. (Distributivity.)
- For all $v_1, v_2 \in V$ and $c \in \mathfrak{R}$ we have $F(cv_1, v_2) = F(v_1, cv_2) = cF(v_1, v_2)$. (Scalar multiplicativity.)
- For all $v \in V$, we have $F(v, v) \geq 0$. Moreover $F(v, v) = 0$ iff $v = 0$.

Notation. We usually do not use a name like F , but write $\langle v, w \rangle$ in place of $F(v, w)$. Often, we also just write $v \cdot w$ and call it a “dot” product.

Warning. Many books will define a more general inner product where the last property of positivity is not assumed in the beginning but later on imposed, because it is essential for definitions of angles and lengths.

Norm, angle. We now use the shortened notation $\langle \cdot, \cdot \rangle$ for an inner product and define

- $\|v\|^2 = \langle v, v \rangle$ or $\|v\| = \sqrt{\langle v, v \rangle}$. This $\|v\|$ is the length of the vector v for the chosen inner product, so strictly speaking, it should carry a marker indicating the inner product. Here, using a function name F helps us put such a marker and write $\|v\|_F$.
- It can be proved that for any two vectors v, w , we have

$$|\langle v, w \rangle| \leq \|v\| \|w\| \text{Cauchy Schwartz Inequality.}$$

Moreover, we get equality iff v, w are linearly dependent.

Further, if v, w are non zero vectors, then $|\langle v, w \rangle| = \|v\| \|w\|$ implies that one of the following two things happens.

Either we have: $\langle v, w \rangle = \|v\| \|w\|$ in case v, w are positive multiples of each other (or can be considered to be in the same direction) or $\langle v, w \rangle = -\|v\| \|w\|$ in case v, w are negative multiples of each other (or can be considered to be in the opposite direction).

- We define the angle between non zero vectors v, w by

$$\angle(v, w) = \arccos\left(\frac{\langle v, w \rangle}{\|v\| \|w\|}\right).$$

The Cauchy Schwartz inequality guarantees that we get a meaningful angle between 0 and 180 degrees.

Warning: One should not lose sight of the fact that this is dependent on the chosen inner product and as before, a marker F can be attached if necessary.

Examples. Here are some examples on inner products in known vector spaces.

- The most common example is in \mathfrak{R}^n . We define $\langle v, w \rangle = v^T w$. This gives the usual dot product. It is obvious that $\|v\|$ corresponds to the usual length of a vector and for $n = 2, 3$, direct calculations can verify the angles to be consistent with usual convention.
- Still in \mathfrak{R}^n a more general inner product can be defined by a **symmetric matrix** $A = A_{n \times n}$ by defining:

$$F \langle v, w \rangle = v^T A w.$$

We may write $\langle v, w \rangle_A$ as a shortened notation, or as an alternative drop all special references to A if no confusion follows.

A random choice of A will not satisfy the positivity condition. It can be shown that a necessary and sufficient condition for a symmetric matrix A to define an inner product is that all its principle minors be positive. This means all the determinants using first few entries of the main diagonal are positive.

- If we go to the space of polynomials P_n or even P , the infinite dimensional space, then we can define an inner product:

$$F(p(t), q(t)) = \int_0^1 p(t)q(t)dt.$$

Clearly, the interval can be changed to other finite intervals leading to different inner products.

- The above example can be generalized to define an inner product on the space $C[a, b]$ which is the space of continuous functions on the interval $[a, b]$. The inner product is defined as

$$F(f(t), g(t)) = \int_a^b f(t)g(t)dt.$$

- In the space of polynomials P_n , define an inner product thus: Choose a set of distinct numbers a_0, a_1, \dots, a_n and define

$$\langle p(t), q(t) \rangle = p(a_0)q(a_0) + p(a_1)q(a_1) + \dots + p(a_n)q(a_n).$$

This defines an inner product. A little thought shows that the map

$$p(t) \rightarrow \begin{bmatrix} p(a_0) \\ p(a_1) \\ \dots \\ p(a_n) \end{bmatrix}$$

is an isomorphism of P_n onto $\mathfrak{R}^{(n+1)}$ and all we are doing is using the usual inner product in the target space $\mathfrak{R}^{(n+1)}$ to define our inner product.

This is a usual method of building new inner products.

Orthogonal sets. Given an inner product $\langle \cdot, \cdot \rangle$ on a vector space V we say that a set of vectors v_1, \dots, v_r is orthogonal, if for any $i \neq j$ we have $\langle v_j, v_i \rangle = 0$. It is easily seen that a set of non zero orthogonal vectors are linearly independent.

Proof. Suppose v_1, \dots, v_r are non zero orthogonal vectors and $c_1v_1 + \dots + c_rv_r = 0$. Take the inner product of both sides with some v_i to get:

$$c_1 \langle v_i, v_1 \rangle + \dots + c_i \langle v_i, v_i \rangle + \dots + c_r \langle v_i, v_r \rangle = \langle v_i, 0 \rangle = 0.$$

Clearly all but the term $c_i \langle v_i, v_i \rangle$ are zero. Moreover, $\langle v_i, v_i \rangle \neq 0$, so $c_i = 0$. Thus each c_i is zero and we have proved independence of our vectors.

This is the most important reason to study and use the inner product!

The set of vectors v_1, \dots, v_r is said to be **orthonormal** if it is orthogonal and also $\langle v_i, v_i \rangle = 1$ for all i . This last condition means that $\|v_i\| = 1$ for each $i = 1, \dots, r$.

Vectors with norm (length) equal to 1 are said to be unit vectors. **Note** that given any non zero vector v , the vector $\pm \frac{v}{\|v\|}$ is always a unit vector. Moreover, if we take the plus sign, then it is in the same direction as v and is in the opposite direction if we use the minus sign.

This gives a simple but useful observation:

Every nonzero vector v is of the form cu where u is a unit vector and $c = \pm\|v\|$.

Coordinate vectors. If we have a set of n non zero orthogonal vectors, v_1, \dots, v_n in an n -dimensional vector space V , then, in view of the above result, they clearly form a basis $B = [v_1 \ v_2 \ \dots \ v_n]$ of V .

Moreover, for any vector $v \in V$, it is easy to find its coordinate vector $[v]_B$ as follows.

Suppose we write $v = c_1v_1 + \dots + c_nv_n$. By taking inner product with v_i and using the same reasoning as above, we see that $\langle v, v_i \rangle = c_i \langle v_i, v_i \rangle$ and thus $c_i = \frac{\langle v, v_i \rangle}{\langle v_i, v_i \rangle}$. This defines the coordinate vector:

$$[v]_B = \begin{bmatrix} c_1 \\ \dots \\ c_n \end{bmatrix} = \begin{bmatrix} \frac{\langle v, v_1 \rangle}{\langle v_1, v_1 \rangle} \\ \dots \\ \frac{\langle v, v_n \rangle}{\langle v_n, v_n \rangle} \end{bmatrix}.$$

Projections. One of the main goals of Linear Algebra is to give efficient methods to solve linear equations $AX = B$. In general, if there are more equations than variables (i.e. A has more rows than columns), then the solutions may not exist. However, in many Scientific and Statistical applications, it makes sense to ask for an answer which makes the equation close to true as much as possible.

If we have an inner product in our vector space, then we can reformulate the problem of solution of $AX = B$ as “find a vector w such that $\|B - Aw\|$ is as small as possible.

This can be shown to be equivalent to finding a w such that $B - Aw$ is orthogonal to each column of A . If we are using the usual inner product in \mathfrak{R}^n , then this is easily seen to be guaranteed by:

$$\text{Normal Equations. . .} \quad A^T Aw = A^T B$$

From the properties of the inner product, we can show that if the columns of A are independent, then the matrix $A^T A$ is invertible. (See proof below). Using this, we get a formal solution:

$$w = (A^T A)^{-1} A^T B.$$

The vector Aw so obtained is geometrically the projection of the vector B into the space $Col A$.

Proof that $A^T A$ is invertible. Suppose if possible, $A^T A$ is singular. Then there is a non zero vector u such that $A^T Au = 0$. Then

$$\langle Au, Au \rangle = u^T A^T Au = u^T (A^T Au) = 0.$$

Hence $Au = 0$. But since columns of A are independent, this implies $u = 0$, a contradiction!

Associated Spaces. Given an $m \times n$ matrix A , we know the two associated spaces $Col A$ and $Nul A$ which are respectively subspaces of \mathfrak{R}^m and \mathfrak{R}^n .

If we use the transpose A^T instead, then we get two other spaces: $Col A^T$ which we call $Row A$ or the row space of A and also $Nul A^T$ or sometimes called the left null space of A .

Note that $Row A$ is a subspace of \mathfrak{R}^n and consists of rows of A transposed into column vectors.

Similarly, $Nul A^T$ is a subspace of \mathfrak{R}^m consisting of all column vectors X such that $A^T X = 0$. Taking transpose, we see that these correspond to row vectors X^T such that $X^T A = 0$. Hence the name of “left null space.”

The concept of inner product gives another meaning to these. Thus, the left null space $Nul A^T$ can be thought of all vectors orthogonal to all vectors of $Col A$.

In general, we define **an orthogonal subspace** to a given space W as $\{v \mid \langle v, w \rangle = 0 \text{ for all } w \in W\}$. We denote this as W^\perp .

It is not hard to see that $(W^\perp)^\perp = W$ for any subspace W . Thus, we note that $Col A = (Nul A^T)^\perp$. This expresses the starting space $Col A$ as a null space of some other matrix. This was the basis of our results

on writing a column space as a null space or conversely, writing a null space as a column space.

Similarly, we can describe $Row A$ as $(Nul A)^\perp$.

It is easy to see that for any subspace W of V we have $\dim(W) + \dim(W^\perp) = \dim(V)$. This is another formulation of the fundamental dimension theorem.

Proof. Write $W = Col A$ for some $m \times n$ matrix A , so that W is a subspace of \mathfrak{R}^m . We know that $\dim(W) = rank(A)$.

Then

$$W^\perp = \{Y \mid \langle w, Y \rangle = 0 \text{ for all } w \in W\}.$$

Since $\langle w, Y \rangle = w^T Y$, we see that $W^\perp = Nul A^T$ and we know that its dimension is $m - rank(A^T) = m - rank(A)$. Thus, we have proved that

$$\dim(W) + \dim(W^\perp) = rank(A) + m - rank(A) = m.$$

Orthonormal Bases.

Suppose that we have a vector space V with an inner product and a given subspace W .

The above results make it clear that we would greatly benefit if given any basis (or even a spanning set) of the subspace W , we can find a suitable orthogonal (or even orthonormal) basis for W from the given set.

This can be accomplished by a slight modification of our row reduction algorithm. This is a way of codifying the Gram-Schmidt process discussed in the book. We show the method below, which is not in the book.

I.P. matrix Suppose that v_1, \dots, v_r is a spanning set for W . First step is to make a matrix M such that $M_{ij} = \langle v_i, v_j \rangle$ for all $i, j = 1, \dots, r$.

Note that M is a symmetric $r \times r$ matrix and we can think of M as $\langle B, B \rangle$ where B is the row of vectors $[v_1 \ v_2 \ \dots \ v_r]$. This is said to be the I.P. (Inner Product) matrix of the spanning set B .

If we replace B by linear combinations of v_1, \dots, v_r then we can think of the new set of vectors as BP where P is the matrix describing the combinations.

If P is invertible, then vectors of BP form a new spanning set for the same space W and its I.P. matrix is $P^T M P$. We shall show that there is an invertible matrix R such that $R^T M R$ is a diagonal matrix.

It follows that the new generating set BR consists of orthogonal vectors.

If the original vectors of B were independent, then the new vectors BR will indeed be an orthogonal basis. Moreover, in this case, the matrix R can be chosen to be (unit) upper triangular. This is known as the Gram-Schmidt theorem.

The Algorithm. Start with the I.P. matrix M of a spanning set v_1, \dots, v_r for W . Let I_r be the usual identity matrix. Set $A = (M|I_r)$ the augmented matrix as usual.

Perform the usual row reductions on M to try and convert it to REF. **However, every time you do a row transformation, immediately follow it with a corresponding column transformation.**

Example 1 Here is an example of three vectors v_1, v_2, v_3 whose I.P. matrix is

$$M = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Note that we did not mention which vector space this is, and the point is that we need not know it! We just need the I.P. matrix for the given spanning set!

The same matrix augmented with Identity matrix is:

$$A = \begin{bmatrix} 2 & 1 & 1 & 1 & 0 & 0 \\ 1 & 5 & 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 & 1 \end{bmatrix}.$$

We make the row transformation $R_2 - (\frac{1}{2})R_1$ and immediately follow it with $C_2 - (\frac{1}{2})C_1$ to get:

$$\begin{bmatrix} 2 & 0 & 1 & 1 & 0 & 0 \\ 0 & 9/2 & 1/2 & -1/2 & 1 & 0 \\ 1 & 1/2 & 2 & 0 & 0 & 1 \end{bmatrix}$$

Now we do $R_3 - (\frac{1}{2})R_1$ followed by $C_3 - (\frac{1}{2})C_1$ to get:

$$\begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 9/2 & 1/2 & -1/2 & 1 & 0 \\ 0 & 1/2 & 3/2 & -1/2 & 0 & 1 \end{bmatrix}.$$

One more pair of steps $R_3 - (\frac{1}{9})R_2$ followed by $C_3 - (\frac{1}{9})C_2$ gives:

$$\begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{9}{2} & 0 & -1/2 & 1 & 0 \\ 0 & 0 & \frac{13}{9} & -4/9 & -1/9 & 1 \end{bmatrix}.$$

The first 3×3 part is now a diagonal matrix and the second 3×3 part is recording the change of basis. Precisely, the second part is R^T , so the new orthogonal basis is given by:

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} 1 & -1/2 & -4/9 \\ 0 & 1 & -1/9 \\ 0 & 0 & 1 \end{bmatrix}.$$

We are thus claiming that $w_1 = v_1$, $w_2 = v_2 - \frac{1}{2}v_1$ and $w_3 = v_3 - \frac{1}{9}v_2 - \frac{4}{9}v_1$ gives a new orthogonal basis for $W = \text{Span}\{v_1, v_2, v_3\}$.

If we need this to be orthonormal, we may further divide each vector by its length. (The lengths are visible in the final matrix as $\sqrt{2}$, $\sqrt{\frac{9}{2}}$ and $\sqrt{\frac{13}{9}}$ respectively!

It is instructive to check that this matches the Gram-Schmidt process.

Here is another example where the starting vectors are not independent.

Example 2 Now let the starting vectors v_1, v_2, v_3 have the following I.P. matrix:

$$M = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 5 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

As before, we make the augmented matrix A :

$$A = \begin{bmatrix} 2 & 1 & 3 & 1 & 0 & 0 \\ 1 & 5 & 6 & 0 & 1 & 0 \\ 3 & 6 & 9 & 0 & 0 & 1 \end{bmatrix}.$$

Perform the following operations and verify the results:

- Perform $R_2 - (\frac{1}{2})R_1$ and $C_2 - (\frac{1}{2})C_1$ to get:

$$\begin{bmatrix} 2 & 0 & 3 & 1 & 0 & 0 \\ 0 & 9/2 & 9/2 & -1/2 & 1 & 0 \\ 3 & 9/2 & 9 & 0 & 0 & 1 \end{bmatrix}.$$

- We give the next two steps without mentioning the operations. Figure out the operations:

$$\begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 9/2 & 9/2 & -1/2 & 1 & 0 \\ 0 & 9/2 & 9/2 & -3/2 & 0 & 1 \end{bmatrix}.$$

$$\begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 9/2 & 0 & -1/2 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{bmatrix}.$$

- Note that the third new vector has norm zero and hence it is zero! the process indicates that the third vector is $w_3 = v_3 - v_2 - v_1 = 0$ and thus it identifies the linear dependence relation too!
- We can now conclude that our vector space $\text{Span}\{v_1, v_2, v_3\}$ is actually two dimensional with $w_1 = v_1$ and $w_2 = v_2 - (\frac{1}{2})v_1$ as an orthogonal basis. The lengths of w_1, w_2 are $\sqrt{2}, \sqrt{\frac{9}{2}}$ respectively.

Summary of topics to study.

1. Calculations of inner products, lengths, angles, unit vectors in given directions.
2. Learn to work with any given inner product(material from 6.7).
3. Orthogonal projection into a subspace. Using the normal equations. Do this when the subspace has an orthogonal basis.
4. Checking for orthogonal or orthonormal vectors.
5. Gram-Schmidt process as described above.
6. (Future work.) Projection into a subspace whose basis may not be orthogonal.
7. (Future work.) Various fitting techniques.

Summary from Chapter 5.

1. Review definitions of eigenvalues, eigenvectors, eigenspaces.
2. Learn how to efficiently calculate the characteristic polynomials.
3. Given bases for eigenspaces of a matrix A , put together a basis for the whole vector space in which the multiplication by A is given by a diagonal matrix. know the condition when this is possible.
4. Diagonalize a given matrix, when possible.
5. Learn a standard form of a 2×2 matrix when it has complex eigenvalues.