MA322
Sathaye **Notes on Inner Products** **Notes on Chapter 6 Spring 2017**

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

- Commutativity: For all $v_1, v_2 \in V$, we have $F(v_1, v_2) = F(v_2, v_1)$.
- Distributivity:For all $v_1, v_2, v_3 \in V$, we have $F(v_1, v_2 + v_3) = F(v_1, v_2) + F(v_1, v_3)$.
- Scalar multiplicativity: For all $v_1, v_2 \in V$ and $c \in \Re$ we have $F(cv_1, v_2) = F(v_1, cv_2) = cF(v_1, v_2)$.
- **Positivity:**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 $< v, w >$ 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 imposed later. The positivity is essential for definitions of angles and lengths.

Norm, angle. We now use the shortened notation $< \ , \ >$ for an inner product and define

- $||v||^2 =< v, v >$ or $||v|| = \sqrt{< v, v >}$. 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

  $$| < v, w > | \leq ||v||||w|| \textbf{Cauchy Schwartz Inequality.}.$$

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

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

  Either we have: $< v, w >= ||v||||w||$ in case v, w are positive multiples of each other (or can be considered to be in the same direction) or $< v, w >= -||v||||w||$ 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

  $$\angle(v, w) = \arccos(\frac{< v, w >}{||v||||w||}.$$

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 of inner products in known vector spaces.

- The most common example is in $\Re^n$. We define $< v, w >= v^T w$. 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 $\Re^n$ a more general inner product can be defined by **a symmetric matrix** $A = A_{n\times n}$ by defining:

  $$F(v, w) = v^T A w.$$

  We may write $< v, w >_A$ 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 $a_0, a_1, \cdots, a_n$ and define

  $$< p(t), q(t) >= p(a_0)q(a_0) + p(a_1)q(a_1) + \cdots + p(a_n)q(a_n).$$

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

  $$p(t) \to \begin{bmatrix} p(a_0) \\ p(a_1) \\ \cdots \\ p(a_n) \end{bmatrix}$$

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

This is a usual method of building new inner products.

Orthogonal sets. Given an inner product $< \ , \ >$ on a vector space V we say that a set of vectors $v_1, \cdots, v_r$ is orthogonal, if for any $i \neq j$ we have $< v_j, v_j >= 0$.

It is easily seen that a set of non zero orthogonal vectors are linearly independent.

Proof. Suppose $v_1, \cdots, v_r$ are non zero orthogonal vectors and $c_1 v_1 + \cdots + c_r v_r = 0$. Take the inner product of both sides with some v_i to get:

$$c_1 < v_i, v_1 > + \cdots + c_i < v_i, v_i > + \cdots + c_r < v_i, v_r > = < v_i, 0 > = 0.$$

Clearly all but the term $c_i < v_i, v_i >$ are zero. Moreover, $< v_i, v_i > \neq 0$, 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 $v_1, \cdots, v_r$ is said to be **orthonormal** if it is orthogonal and also $< v_i, v_i >= 1$ for all i. This last condition means that $||v_i|| = 1$ for each $i = 1, \cdots r$.

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, $v_1, \cdots, v_n$ in an n-dimensional vector space V, then, in view of the above result, they clearly form a basis $B = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ of V.

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

Suppose we write $v = c_1 v_1 + \cdots c_n v_n$. By taking inner product with v_i and using the same reasoning as above, we see that $< v, v_i >= c_i < v_i, v_i >$ and thus $c_i = \frac{< v, v_i >}{< v_i, v_i >}$. This defines the coordinate vector:

$$[v]_B = \begin{bmatrix} c_1 \\ \cdots \\ c_n \end{bmatrix} = \begin{bmatrix} \frac{< v, v_1 >}{< v_1, v_1 >} \\ \cdots \\ \frac{< v, v_n >}{< v_n, v_n >} \end{bmatrix}.$$

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 $\Re^n$, then this is easily seen to be guaranteed by:

Normal Equations. $$A^T Aw = A^T B$$

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 $\mathbf{A^T A}$ 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

$$< Au, Au >= u^T A^T Au = u^T (A^T Au) = 0.$$

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 $\Re^m$ and $\Re^n$.

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 $\Re^n$ and consists of rows of A **transposed into column vectors.**

Similarly, $Nul(A^T)$ is a subspace of $\Re^m$ consisting of all column vectors X such that $A^T X = 0$. Taking transpose, we see that these correspond to row vectors X 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 $\{v \mid <v, w>= 0 \text{ for all } w \in W\}$. 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.

Note We already know another method to find this left null space. Recall the consistency matrix G obtained by finding an REF of $(A|I)$ and taking the part of the transformed I in front of zero rows in REF of A. We know that vectors $v \in Col(A)$ are characterized by $Gv = 0$. This means $v^T G^T = 0$ and thus $Nul(A^T) = Col(G^T))$ as desired.

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 $\Re^m$. We know that $\dim(W) = rank(A)$.

Then
$$W^\perp = \{Y \mid <w, Y>= 0 \text{ for all } w \in W\}.$$

Since $<w, Y>= w^T Y$, 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 $v_1, \cdots v_r$ is a spanning set for W. First step is to make a matrix M^* such that $M^*_{ij} =< v_i, v_j >$ for all $i, j = 1, \cdots r$.

Note that M^* is a symmetric $r \times r$ matrix and we can think of M^* as $< B, B >$ where B is the row of vectors $\begin{bmatrix} v_1 & v_2 & \cdots & v_r \end{bmatrix}$. This is said to be the I.P. (Inner Product) matrix of the spanning set B.

If we replace B by linear combinations of $v_1, \cdots v_r$ 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. We present a different Gram-Schmidt algorithm which is lot easier to implement than the one in the book. It is based on just the usual row reduction algorithm and does not involve complicated expressions.

Start with the I.P. matrix M^* of a spanning set $v_1, \cdots, v_r$ for W. If we are working with the "usual inner product" in $\Re^n$, then we simply have $M^* = A^T A$ where A is the matrix whose columns are $v_1, \cdots, v_r$. In either case, M^* is a square $r \times r$ square matrix.

Let I_r be the usual identity matrix. Set $M = (M^* | I_r)$ the augmented matrix as usual.

Perform the usual row reductions on M to try and convert it to REF to get a matrix $G = (N|S)$ where N is upper triangular and S is the lower triangular, invertible matrix such that $G = SM^*$.

(Note that for N to be upper triangular one can only do row operations in which a row is modified by adding to it a multiple of row above it and in which a row with a zero in the intended pivot position is interchanged with one below it. Scaling rows is permitted.)

Let $R = S^T$. Since R and $R^T M^* = SM^*$ are both square, upper triangular, so is their product $R^T M^* R$.

In case $M^* = A^T A$ this expression is $R^T A^T A R =< AR, AR >$.

In the more general case with $M^* =< B, B >$ we get that $R^T M^* R =< BR, BR >$.

Finally, we note that this matrix $R^T M^* R$ is upper triangular and equal to its own transpose, hence it must be a diiagonal matrix!

It follows immediately that the columns of AR (or vectors in BR)are mutually orthogonal.

If columns of A are linearly dependent, then we will simply have some columns in AR as zero columns and the non zero columns will give a basis of $Col(A)$.

Similarly, if $v_1, \cdots, v_r$ are linearly dependent, then some of the vectors in BR will be zero and we get a basis for $Span(B)$ after dropping them.

Simple Example Let $A = \begin{bmatrix} -2 & 1 \\ 2 & -2 \\ 1 & 1 \end{bmatrix}$. Then the innerproduct matrix

$$IP = A^t A = \begin{bmatrix} 9 & -5 \\ -5 & 6 \end{bmatrix}.$$

We augment IP by the identity to get

$$M = \left[\begin{array}{cc|cc} 9 & -5 & 1 & 0 \\ -5 & 6 & 0 & 1 \end{array}\right]$$

The row operation $R_2 \to R_2 + \frac{5}{9} R_1$ produces $\left[\begin{array}{cc|cc} 9 & -5 & 1 & 0 \\ 0 & \frac{29}{9} & \frac{5}{9} & 1 \end{array}\right]$

We stop at this point since the IP matrix is now in REF.

Now we take R to be the transpose of the matrix derived from the identity. $R = \begin{bmatrix} 1 & \frac{5}{9} \\ 0 & 1 \end{bmatrix}$

Then the row operations have multiplied $A^T A$ by R^T on the left and the column operations have multiplied it by R so we have $R^T(A^T A)R = \begin{bmatrix} 9 & 0 \\ 0 & \frac{29}{9} \end{bmatrix}$.

This says that $(AR)^t(AR) = \begin{bmatrix} 9 & 0 \\ 0 & \frac{29}{9} \end{bmatrix}$

Since R is invertible, AR has the same linear span as the columns of A. So the columns of AR are spanning set for $col(A)$. Since they are mutually independent they are then an independent spanning set for $Col(A)$ so the columns of

$$AR = \begin{bmatrix} -2 & 1 \\ 2 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{5}{9} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -\frac{1}{9} \\ 2 & -\frac{8}{9} \\ 1 & \frac{14}{9} \end{bmatrix}.$$

are an orthoginal basis for the column space of A

If we want an orthonormal basis then since $(AR)^T(AR) = \begin{bmatrix} 9 & 0 \\ 0 & \frac{29}{9} \end{bmatrix}$ we have that 9 is the square of the length of column 1 of AR and $\frac{9}{29}$ is the square of the length of column 2. To convert an orthogonal set to an orthonormal set all we have to do is divide each element by its length. That is we divide column 1 by 3 and column 2 by $\sqrt{\frac{29}{9}}$ to get $S = \begin{bmatrix} -\frac{2}{3} & -\frac{\sqrt{29}}{87} \\ \frac{2}{3} & -8\frac{\sqrt{29}}{87} \\ \frac{1}{3} & 14\frac{\sqrt{29}}{87} \end{bmatrix}$. which is a matrix whose columns are an orthonormal basis for $Col(A)$. We can check the orthonormality by calculating $S^T S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

In the above example the matrix A had independent columns. Even if A has dependent columns then the matrix AR will still have mutually orthogonal columns- only some of them would be the zero vector (which is orthogonal to everything). Then all one has to do is delete the zero columns to have an orthogonal basis.

Example 1 Here is an example where the starting vectors are not independent. To emphasize the basic details we give only the IP matrix.

Suppose the matrix A has columns v_1, v_2, v_3 and the following I.P. matrix:

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

As before, we make the augmented matrix $< A^TA|I >$:

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 3 & 1 & 0 & 0 \\ 1 & 5 & 6 & 0 & 1 & 0 \\ 3 & 6 & 9 & 0 & 0 & 1 \end{array}\right].$$

- $$\left[\begin{array}{ccc|ccc} 2 & 1 & 3 & 1 & 0 & 0 \\ 0 & 9/2 & 9/2 & -1/2 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right].$$

- $$R^T = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}.$$

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

$$R^TA^TAR = (AR)^T(AR) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 9/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- Note that the third new vector (column 3 of AR) has norm zero and hence it is zero! This is $A \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ so 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 $Span\{v_1, v_2, v_3\}$ 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.