HW: In $\Re^n$, why $\|\mathbf{u}\|_p$ may not have an inner product for $p \neq 2$?

Motivation:

- Consider the following inner product on $\Re^2$: For any $\mathbf{x}, \mathbf{y} \in \Re^2$, let $< \mathbf{x}, \mathbf{y} >= 2x_1y_1 - x_1y_2 - x_2y_1 + 4x_2y_2$. It can be easily verified that this in an inner product (by checking for linearity, symmetry and positive definiteness by expressing it as a sum of squares).
- This inner product is certainly different from the conventional (Eucledian) dot product $< \mathbf{x}, \mathbf{y} >_E = x_1y_1 + x_2y_2$ which corredponds to the $\|.\|_2$ norm.
- Is it possible that the $< \mathbf{x}, \mathbf{y} >$ defined in step 1 (or some other such inner product) corresponds to $\|.\|_p$ norm for $p \neq 2$?

HW: In $\Re^n$, why $\|\mathbf{u}\|_p$ may not have an inner product for $p \neq 2$?

Motivation:

- Consider the following inner product on $\Re^2$: For any $\mathbf{x}, \mathbf{y} \in \Re^2$, let $< \mathbf{x}, \mathbf{y} >= 2x_1y_1 - x_1y_2 - x_2y_1 + 4x_2y_2$. It can be easily verified that this in an inner product (by checking for linearity, symmetry and positive definiteness by expressing it as a sum of squares).
- This inner product is certainly different from the conventional (Eucledian) dot product $< \mathbf{x}, \mathbf{y} >_E = x_1y_1 + x_2y_2$ which corredponds to the $\|.\|_2$ norm.
- Is it possible that the $< \mathbf{x}, \mathbf{y} >$ defined in step 1 (or some other such inner product) corresponds to $\|.\|_p$ norm for $p \neq 2$?

In $\Re^n$, it can be proved that for any inner product vector space $(\mathcal{V}, < .,. >)$, the inner product $< .,. >$ (including the Eucledian one) can be represented as

$$< \mathbf{u}, \mathbf{v} >= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j < \mathbf{e}_i, \mathbf{e}_j >= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{a}^T E \mathbf{b} =< \mathbf{a}^T, \mathbf{b} >_E$$

Recap: Basis and Dimensions from Linear Algebra wrt $< .,. >_E$ (Eucledian Inner Product) (For your homework)

Instructor: Prof. Ganesh Ramakrishnan

Recap: Basis and Dimensions from Linear Algebra (For your homework)

Recap: Basis in Linear Algebra

Basis for a space: *The basis for a space is a set of vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *with two properties, viz., (1) The vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *are independent and (2) These vectors span the space.*

Set of vectors that is necessary and sufficient for spanning the space.
Eg: A (standard) basis for the four dimensional space $\Re^4$ is:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (3)$$

It is easy to verify that the above vectors are independent; if a combination of the vectors using the scalars in $[c_1, c_2, c_3, c_4]$ should yield the zero vector, we must have $c_1 = c_2 = c_3 = c_4 = 0$. Another way of proving this is by making the four vectors the columns of a matrix. The resultant matrix will be an identity matrix. The null space of an identity matrix is the zero vector.

Recap: Basis in Linear Algebra (contd.)

This is not the only basis of $\Re^4$. Consider the following three vectors

$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \end{bmatrix} \quad (4)$$

These vectors are certainly independent. But they do not span $\Re^4$.
This can be proved by showing that the following vector in $\Re^4$ cannot be expressed as a linear combination of these vectors.

$$\begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \end{bmatrix} \quad (5)$$

Recap: Basis in Linear Algebra (contd.)

- In fact, if the last vector on the previous slide is added to the set of three vectors in (4), together, they define another basis for $\Re^4$.
- This could be proved by introducing them as columns of a matrix A, subject A to row reduction and check if there are any free variables (or equivalently, whether all columns are pivot columns). If there are no free variables, we can conclude that the vectors form a basis for $\Re^4$.
- This is also equivalent to the statement that *if the matrix A is invertible, its columns form a basis for its column space*.

Recap: Basis in Linear Algebra (contd.)

- We can generalize our observations to $\Re^n$: *if an $n \times n$ matrix A is invertible, its coulumns for a basis for $\Re^n$.*
- While there can be many bases for a space, a commonality between all the bases is that they have exactly the same number of vectors.
- This unique size of the basis is called the dimension of the space.

Dimension: *The number of vectors in any basis of a vector space is called the dimension of the space.*

Recap: Basis in Linear Algebra (contd.)

Do the vectors in (4), form a basis for any space at all?
The vectors are independent and therefore span the space of all linear combinations of the three vectors.
The space spanned by these vectors is a hyperplane in $\Re^4$.
Let A be any matrix. By definition, the columns of A span the column space $C(A)$ of A. If there exists a $\mathbf{c} \neq \mathbf{0}$ such that, $A\mathbf{c} = 0$, then the columns of A are not linearly independent.
For example, the columns of the matrix A given below are not linearly independent.

$$A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & 3 & 5 & 2 \\ 3 & 4 & 7 & 3 \end{bmatrix} \tag{6}$$

A choice of $\mathbf{c} = [-1\ 0\ 0\ 1]^T$ gives $A\mathbf{c} = 0$. Thus, the columns of A do not form a basis for its columns space.

Recap: Basis in Linear Algebra (contd.)

What is a basis for $C(A)$? A most natural choice is the first two columns of A; the thid column is the sum of the first and second columns, while the fourth column is the same as the first column. Also, column elimination[1] on A yields pivots on the first two columns. Thus, a basis for $C(A)$ is

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \quad (7)$$

Another basis for $C(A)$ consists of the first and third columns. We note that the dimension of $C(A)$ is 2. We also note that the rank of A is the number of its pivots columns, which is exactly the dimension of $C(A)$.

[1] Column elimination operations are very similar to row elimination operations.

Recap: Basis in Linear Algebra (contd.)

All of this gives us a nice result.

Theorem

The rank of a matrix is the same as the dimension of its column space. That is, $rank(A) = dimension\left(C(A)\right)$.

- What about the dimension of the null space? We already saw that $\mathbf{c} = [-1\ 0\ 0\ 1]^T$ is in the null space.
- Another element of the null space is $\mathbf{c}' = [1\ 1\ \ -1\ 0]^T$. These vectors in the null space specify combinations of the columns that yield zeroes. The two vectors $\mathbf{c}$ and $\mathbf{c}'$ are obviously independent. Do these two vectors span the entire null space?
- The dimension of the null space is the same as the number of free variables, which happens to be $4 - 2 = 2$ in this example. Thus the two vectors c and c' must indeed span the null space. In fact, it can be proved that the dimension of the null space of an $m \times n$ matrix A is $n - rank(A)$.

Recap: Row Space and Column Space in Linear Algebra (contd.)

- The space spanned by the rows of a matrix is called the *row space*. We can also define the row space of a matrix A as the column space of its transpose A^T. Thus the row space of A can be specified as $C(A^T)$.
- The null space of A, $N(A)$ is often called the *right null space* of A, while the null space of A^T, $N(A^T)$ is often referred to as its *left null space*.
- How do we visualize these four spaces? $N(A)$ and $C(A^T)$ of an $m \times n$ matrix A are in $\Re^n$, while $C(A)$ and $N(A^T)$ are in $\Re^m$.
- How can we construct bases for each of the four subspaces? We note that dimensions of $C(A)$ and the rank of $C(A^T)$ should be the same, since row rank of a matrix is its column rank. The bases of $C(A)$ can be obtained as the set of the pivot columns.

Recap: The Four Subspaces and their Bases (contd.)

- Let r be the rank of A. Recall that the null space is constructed by linear combinations of the special solutions of the null space (**??**) and there is one special solution for each assignment of the free variables. In fact, the number of special solutions exactly equals the number of free variables, which is $n - r$. Thus, the dimension of $N(A)$ will be $n - r$.
- Similarly, the dimension of $N(A^T)$ will be $m - r$.

Let us illustrate all this on the sample matrix in (6).

$$\begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & 3 & 5 & 2 \\ 3 & 4 & 7 & 3 \end{bmatrix} \overset{E_{2,1},E_{3,1}}{\Longrightarrow} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & -2 & -2 & 0 \end{bmatrix} \overset{E_{3,2}}{\Longrightarrow} (R =) \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad (8)$$

Recap: The Four Subspaces and their Bases (contd.)

- The reduced matrix R has the same row space as A, by virtue of the nature of row reduction. In fact, the rows of A can be retrieved from the rows of R by reversing the linear operations involved in row elimination. The first two rows give a basis for the row space of A.
- The dimension of $C(A^T)$ is 2, which is also the rank of A.
- To find the left null space of A, we look at the system $\mathbf{y}^T A = 0$. Recall the Gauss-Jordan elimination method from Section **??** that augments A with an $m \times m$ identity matrix, and performs row elimination on the augmented matrix.

$$[A\ I_{m\times m}] \overset{rref}{\Longrightarrow} [R\ E_{m\times m}]$$

The rref will consist of the reduced matrix augmented with the elimination matrix reproduced on its right.

Recap: The Four Subspaces and their Bases (contd.)

For the example case in 8, we apply the same elimination steps to obtain the matrix E below:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \stackrel{E_{2,1},E_{3,1}}{\Longrightarrow} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \stackrel{E_{3,2}}{\Longrightarrow} (E =) \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \tag{9}$$

Recap: The Four Subspaces and their Bases (contd.)

Writing down $EA = R$,

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & 3 & 5 & 2 \\ 3 & 4 & 7 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{10}$$

We observe that the last row of E specifies a linear combination of the rows of A that yields a zero vector (corresponding to the last row of R). This is the only vector that yields a zero row in R and is therefore the only element in the basis of the left null space of A, that is, $N(A^T)$. The dimension of $N(A^T)$ is 1.

Recap: The Four Subspaces and their Bases (contd.)

- As another example, consider the space $\mathcal{S}$ of vectors $\mathbf{v} \in \Re^3$ where $\mathbf{v} = [v_1 \ v_2 \ v_3]^T$ such that $\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$. What is the dimension of this subspace?
- Note that this subspace is the right null space $N(A)$ of a 1×3 matrix $A = [1 \ 1 \ 1]$, since $A\mathbf{v} = 0$. The rank, $r = rank(A)$ is 1, implying that the dimension of the right null space is $n - r = 3 - 1 = 2$.
- One set of basis vectors for $\mathcal{S}$ is $[-1 \ 1 \ 0]$, $[-1 \ 0 \ 1]$. The column space $C(A)$ is $\Re^1$ with dimension 1. The left null space $N(A^T)$ is the singleton set $\{0\}$ and as expected, has a dimension of $m - r = 1 - 1 = 0$.

Recap: Matrix Spaces

We will extend the set of examples of vector spaces discussed in Section **??** with a new vector space, that of all $m \times n$ matrices with real entries, denoted by $\Re^{m \times n}$.

- It is easy to verify that the space of all matrices is closed under operations of addition and scalar multiplication. Additionally, there are interesting subspaces in the entire matrix space $\Re^{m \times n}$, *viz.*,
  - set $\mathcal{S}$ of all $n \times n$ symmetric matrices
  - set $\mathcal{U}$ of all $n \times n$ upper triangular matrices
  - set $\mathcal{L}$ of all $n \times n$ lower triangular matrices
  - set $\mathcal{D}$ of all $n \times n$ diagonal matrices
- Let $\mathcal{M} = \Re^{3 \times 3}$ be the space of all 3×3 matrices. The dimension of $\mathcal{M}$ is 9. Each element of this basis has a 1 in one of the 9 positions and the remaining entries as zeroes.
- Of these basis elements, three are symmetric (those having a 1 in any of the diagonal positions). These three matrices form the basis for the subspace of diagonal matrices.
- Six of the nine basis elements of M form the basis of $\mathcal{U}$ while six of them form the basis of $\mathcal{L}$.

Recap: Matrix Spaces (contd.)

- The intersection of any two matrix spaces is also a matrix space. For example, $\mathcal{S} \cap \mathcal{U}$ is $\mathcal{D}$, the set of diagonal matrices.
- However the union of any two matrix spaces need not be a matrix space. For example, $\mathcal{S} \cup \mathcal{U}$ is not a matrix space; the sum $S + U$, $S \in \mathcal{S}$, $U \in \mathcal{U}$ need not belong to $\mathcal{S} \cup \mathcal{U}$.
- We will discuss a special set comprising all linear combinations of the elements of union of two vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ (*i.e.*, $\mathcal{V}_1 \cup \mathcal{V}_2$), and denote this set by $\mathcal{V}_1 \oplus \mathcal{V}_2$. By definition, this set is a vector space. For example, $\mathcal{S} + \mathcal{U} = \mathcal{M}$, which is a vector space.

Recap: Matrix Spaces (contd.)

A property fundamental to many properties of matrices is the expression for a rank 1 matrix. A rank 1 matrix can be expressed as the product of a column vector with a row vector (the row vector forming a basis for the matrix). Thus, any rank 1 matrix X can be expressed as

$$X_{m\times n} = u^T v = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ . \\ . \\ u_m \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \tag{11}$$

Recap: Matrix Spaces (contd.)

Let $\mathcal{M}_{m\times n}$ be the set of all $m \times n$ matrices. Is the subset of $\mathcal{M}_{m\times n}$ matrices with rank k, a subspace? For $k = 1$, this space is obviously not a vector space as is evident from the sum of rank 1 matrices, A^1 and B^1, which is not a rank 1 matrix. In fact, the subset of $\mathcal{M}_{m\times n}$ matrices with rank k is not a subspace.

$$A^1 + B^1 = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 2 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 4 & 2 \\ 2 & 2 & 1 \\ 4 & 4 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 6 & 3 \\ 4 & 6 & 2 \\ 5 & 6 & 3 \end{bmatrix} \quad (12)$$

Orthogonality and Projection

- Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be orthogonal *iff*, their dot product (more generally, the inner product) is 0. In the eucledian space, the dot product of the two vectors is $x^T y$.
- The condition $x^T y = 0$ is equivalent to the pythagorous condition between the vectors $\mathbf{x}$ and $\mathbf{y}$ that form the perpendicular sides of a right triangle with the hypotenuse given by $\mathbf{x} + \mathbf{y}$. The *pythagorous condition* is $||\mathbf{x}||^2 + ||\mathbf{y}||^2 = ||\mathbf{x} + \mathbf{y}||^2$, where the norm is the eucledian norm, given by $||\mathbf{x}||^2 = \mathbf{x}^T\mathbf{x}$.
- This equivalence can be easily proved and is left to the reader as an exercise. By definition, the vector $\mathbf{0}$ is orthogonal to every other vector.

Orthogonality and Projection

We will extend the definition of orthogonality to subspaces; a subspace $\mathcal{U}$ is orthogonal to subspace $\mathcal{V}$ *iff*, every vector in $\mathcal{U}$ is orthogonal to every vector in $\mathcal{V}$. As an example:

Theorem

The row space $C(A^T)$ of an $m \times n$ matrix A is orthogonal to its right null space $N(A)$.

Proof: $A\mathbf{x} = \mathbf{0}, \ \forall \mathbf{x} \in N(A)$. On the other hand, $\forall \ \mathbf{y} \in C(A^T), \ \exists \ z \in \Re^m, \ \mathbf{s}.t., \ \mathbf{y} = A^T\mathbf{z}$.
Therefore, $\forall \ \mathbf{y} \in C(A^T), \ \mathbf{x} \in N(A), \mathbf{y}^T\mathbf{x} = \mathbf{z}^T A\mathbf{x} = \mathbf{z}.\mathbf{0} = 0$. □

Orthogonality and Projection

Not only are $C(A^T)$ and the right null space $N(A)$ orthogonal to each other, but they are also *orthogonal complements* in $\Re^n$, that is, $N(A)$ contains all vectors that are orthogonal to some vector in $C(A^T)$.

Theorem

The null space of A and its row space are orthogonal complements.

Proof: We note, based on our discussion earlier that the dimensions of the row space and the (right) null space add up to n, which is the number of columns of A. For any vector $\mathbf{y} \in C(A^T)$, we have $\exists\ \mathbf{z} \in \Re^m,\ \mathbf{s}.t.,\ \mathbf{y} = A^T\mathbf{z}$. Suppose $\forall\ \mathbf{y} \in C(A^T),\ \mathbf{y}^T\mathbf{x} = 0$. That is, $\forall\ \mathbf{z} \in \Re^m,\ \mathbf{z}^T A\mathbf{x} = 0$. This is possible only if $A\mathbf{x} = \mathbf{0}$. Thus, necessarily, $\mathbf{x} \in N(A)$. □

Along similar lines, we could prove that the column space $C(A)$ and the left null space $N(A^T)$ are orthogonal complements in $\Re^m$.

Orthogonality and Projection

Based on preceeding theorem,we prove that there is a one-to-one mapping between the elements of row space and column space.

Theorem

If $\mathbf{x} \in C(A^T)$, $\mathbf{y} \in C(A^T)$ *and* $\mathbf{x} \neq \mathbf{y}$, *then,* $A\mathbf{x} \neq A\mathbf{y}$.

Proof: Note that $A\mathbf{x}$ and $A\mathbf{y}$ are both elements of $C(A)$. Next, observe that $\mathbf{x} - \mathbf{y} \in C(A^T)$, which by theorem 6, implies that $\mathbf{x} - \mathbf{y} \notin N(A)$. Therefore, $A\mathbf{x} - A\mathbf{y} \neq \mathbf{0}$ or in other words, $A\mathbf{x} \neq A\mathbf{y}$. □

Similarly, it can be proved that if $\mathbf{x} \in C(A)$, $\mathbf{y} \in C(A)$ and $\mathbf{x} \neq \mathbf{y}$, then, $A^T\mathbf{x} \neq A^T\mathbf{y}$. The two properties together imply a one-to-one mapping between the row and column spaces.

Projection Matrices

The projection of a vector $\mathbf{t}$ on a vector $\mathbf{s}$ is a vector $\mathbf{p} = c\mathbf{s}$, $c \in \Re$ (in the same direction as $\mathbf{s}$), such that $\mathbf{t} - c\mathbf{s}$ is orthogonal to $\mathbf{s}$. That is, $\mathbf{s}^T(\mathbf{t} - c\mathbf{s}) = 0$ or $\mathbf{s}^T\mathbf{t} = c\mathbf{s}^T\mathbf{s})$. Thus, the scaling factor c is given by $c = \frac{\mathbf{s}^T\mathbf{t}}{\mathbf{s}^T\mathbf{s}}$. The projection of the vector $\mathbf{t}$ on a vector $\mathbf{s}$ is then

$$\mathbf{p} = \mathbf{s}\frac{\mathbf{t}^T\mathbf{s}}{\mathbf{s}^T\mathbf{s}} \tag{13}$$

Using the associative property of matrix multiplication, the expression for $\mathbf{p}$ can be re-written as

$$\mathbf{p} = P\mathbf{t} \tag{14}$$

where, $P = \mathbf{s}\mathbf{s}^T\frac{1}{\mathbf{s}^T\mathbf{s}}$ is called the *projection matrix*.

Projection Matrices (contd.)

- The rank of the projection matrix is 1 (since it is a column mutiplied by a row).
- The projection matrix is symmetric and its column space is a line through s.
- For any $d \in \Re$, $P(d\mathbf{s}) = d\mathbf{s}$, that is, the projection of any vector in the direction of $\mathbf{s}$ is the same vector. Thus, $P^2 = P$.

Least Squares

- We earlier saw a method for solving the system $A\mathbf{x} = \mathbf{b}$ (A being an $m \times n$ matrix), when a solution exists. Howevever, a solution may not exist, especially when $m > n$, that is when the number of equations is greater than the number of variables.
- We also saw that the *rref* looks like $[I\ \mathbf{0}]^T$, where I is an $n \times n$ identity matrix. It could happen that the row reduction yields a zero submatrix in the lower part of A, but the corresponding elements in $\mathbf{b}$ are not zeroes.
- In other words, $\mathbf{b}$ may not be in the column space of A. In such cases, we are often interested in finding a 'best fit' for the system; a solution $\hat{x}$ that satisfies $A\mathbf{x} = \mathbf{b}$ as well as possible.

Projection Matrices (contd.)

- We define the best fit in terms of a vector $\mathbf{p}$ which is the projection of $\mathbf{b}$ onto $C(A)$ and solve $A\hat{\mathbf{x}} = \mathbf{p}$. We require that $\mathbf{b} - \mathbf{p}$ is orthogonal to $C(A)$, which means

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \tag{15}$$

- The vector $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$ is the error vector and is in $N(A^T)$. The equation (59) can be rewritten as

$$(A^TA)\hat{\mathbf{x}} = A^T\mathbf{b} \tag{16}$$

Projection Matrices (contd.)

A matrix that plays a key role in this problem is A^TA. It is an $n \times n$ symmetric matrix (since $(A^TA)^T = A^TA$). The right null space $N(A^TA)$ is the same as $N(A)^2$. It naturally follows that the ranks of A^TA and A are the same (since, the sum of the rank and dimension of null space equal n in either case). Thus, A^TA is invertible exactly if $N(A)$ has dimension 0, or equivalently, A is a full column rank.

Theorem

If A is a full column rank matrix (that is, its columns are independent), A^TA is invertible.

Proof: We will show that the null space of A^TA is $\{0\}$, which implies that the square matrix A^TA is full column (as well as row) rank is invertible. That is, if $A^TA\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = \mathbf{0}$. Note that if $A^TA\mathbf{x} = \mathbf{0}$, then $\mathbf{x}^TA^TA\mathbf{x} = ||A\mathbf{x}|| = 0$ which implies that $A\mathbf{x} = \mathbf{0}$. Since the columns of A are linearly independent, its null space is $\mathbf{0}$ and therefore, $\mathbf{x} = \mathbf{0}$. □

Assuming that A is full column rank, the equation (16) can be rewritten as

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}. \tag{17}$$

Projection Matrices (contd.)

Therefore the expression for the projection $\mathbf{p}$ will be

$$\mathbf{p} = A(A^T A)^{-1} A^T \mathbf{b} \tag{18}$$

This expression is the n-dimensional equivalent of the one dimensional expression for projection in (56). The projection matrix in (18) is given by $P = A(A^T A)^{-1} A^T$.
We will list the solution for some special cases:

- If A is an $n \times n$ square invertible matrix, its column space is the entire $\Re^n$ and the projection matrix will turn out to be the identity matrix.
- Also, if b is in the column space $C(A)$, then $\mathbf{b} = A\mathbf{t}$ for some $t\ in \Re^n$ and consequently, $P\mathbf{b} = A(A^T A)^{-1}(A^T A)\mathbf{t} = A\mathbf{t} = \mathbf{b}$.
- On the other hand, if b is orthogonal to $C(A)$, it will lie in $N(A^T)$, and therefore, $A^T\mathbf{b} = 0$, implying that $\mathbf{p} = 0$.

Projection Matrices (contd.)

Another equivalent way of looking at the best fit solution $\hat{\mathbf{x}}$ is a solution that minimizes the square of the norm of the error vector

$$e(\hat{\mathbf{x}}) = ||A\mathbf{x} - \mathbf{b}||^2 \tag{19}$$

Setting $\frac{de(\hat{\mathbf{x}})}{d\mathbf{x}} = 0$, we get the same expression for $\hat{\mathbf{x}}$ as in (60). The solution in 60 is therefore often called the *least squares solution*. Thus, we saw two views of finding a best fit; first was the view of projecting into the column space while the second concerned itself with minimizing the norm squared of the error vector.

Projection Matrices (contd.)

We will take an example. Consider the data matrix A and the coefficient matrix $\mathbf{b}$ as in (20).

$$Ax = \begin{bmatrix} 2 & -1 \\ -1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} \tag{20}$$

Projection Matrices (contd.)

The matrix A is full column rank and therefore A^TA will be invertible. The matrix A^TA is given as

$$A^TA = \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix}$$

Substituting the value of A^TA in the system of equations (16), we get,

$$6\hat{x}_1 - 3\hat{x}_2 = 2 \quad (21)$$

$$-3\hat{x}_1 + 6\hat{x}_2 = 8 \quad (22)$$

The solution of which is, $x_1 = \frac{4}{5}$, $x_2 = \frac{26}{15}$.

Orthonormal Vectors

A collection of vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ is said to be orthonormal *iff* the following condition holds $\forall\ i, j$:

$$\mathbf{q}_i^T \mathbf{q}_j \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \tag{23}$$

A large part of numerical linear algebra is built around working with orthonormal matrices, since they do not overflow or underflow. Let Q be a matrix comprising the columns $\mathbf{q}_1$ through $\mathbf{q}_n$. It can be easily shown that

$$Q^T Q = I_{n \times n}$$

Orthonormal Vectors (contd.)

When Q is square, $Q^{-1} = Q^T$. Some examples of matrices with orthonormal columns are:

$$Q_{rotation} = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix}, \; Q_{reflection} = \begin{bmatrix} cos(\theta) & sin(\theta) \\ sin(\theta) & -cos(\theta) \end{bmatrix},$$

$$Q_{Hadamard} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \; Q_{rect} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{24}$$

The matrix $Q_{rotation}$ when multiplied to a vector, rotates it by an angle θ, whereas $Q_{reflection}$ reflects the vector at an angle of $\theta/2$.

Orthonormal Vectors (contd.)

These matrices present standard varieties of linear transformation, but in general, premultiplication by an $m \times n$ matrix transforms from an input space in $\Re^m$ to an input space in $\Re^n$.

The matrix $Q_{Hadamard}$ is an orthonormal matrix consisting of only 1's and -1's. Matrices of this form exist only for specific dimensions such as 2, 4, 8, 16, *etc.*, and are called *Hadamard matrices*[3].

The matrix Q_{rect} is an example rectangular matrix whose columns are orthonormal.

[3] An exhaustive listing of different types of matrices can be found at `http://en.wikipedia.org/wiki/List_of_matrices`.

Orthonormal Vectors (contd.)

Suppose a matrix Q has orthonormal columns. What happens when we project any vector onto the column space of Q? Substituting $A = Q$ in (18), we get[4]:

$$\mathbf{p} = Q(Q^TQ)^{-1}Q^T\mathbf{b} = QQ^T\mathbf{b} \tag{25}$$

Making the same substitution in (60),

$$\hat{\mathbf{x}} = (A^TQ)^{-1}Q^T\mathbf{b} = Q^T\mathbf{b} \tag{26}$$

The i^{th} component of $\mathbf{x}$, is given by $x_i = q_i^Tb$.

Let Q_1 be one orthonormal basis and Q_2 be another orthonormal basis for the same space. Let A be the coefficient matrix for a set of points represented using Q_1 and B be the coefficient matrix for the same set of points represented using Q_2. Then $Q_1A = Q_2B$, which implies that B can be computed as $B = Q_2^TQ_1A$. This gives us the formula for changing basis.

[4] Note that $Q^TQ = I$. However, $QQ^T = I$ only if Q is a square matrix.

Gram-Schmidt Orthonormalization

- The goal of the Gram-Schmidt orthonormalization process is to generate a set of orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$, given a set of independent vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$.
- The first step in this process is to generate a set of orthogonal vectors $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n$ from $\mathbf{a}_1, \mathbf{a}_2, \ldots, \bullet\, a_n$. To start with, $\mathbf{t}_1$ is chosen to be $\mathbf{a}_1$.
- Next, the vector $\mathbf{t}_2$ is obtained by removing the projection of $\mathbf{a}_2$ on $\mathbf{t}_1$, from $\mathbf{a}_2$, based on (56). That is,

$$\mathbf{t}_2 = \mathbf{a}_2 - \frac{1}{\mathbf{a}_1^T \mathbf{a}_1} \mathbf{a}_1 \mathbf{a}_1^T \mathbf{a}_2 \tag{27}$$

- This is carried out iteratively for $i = 1, 2, \ldots, n$, using the expression below:

$$\mathbf{t}_i = \mathbf{a}_i - \frac{1}{\mathbf{t}_1^T \mathbf{t}_1} \mathbf{t}_1 \mathbf{t}_1^T \mathbf{a}_i - \frac{1}{\mathbf{t}_2^T \mathbf{t}_2} \mathbf{t}_2 \mathbf{t}_2^T \mathbf{a}_i - \ldots - \frac{1}{\mathbf{t}_{i-1}^T \mathbf{t}_{i-1}} \mathbf{t}_{i-1} \mathbf{t}_{i-1}^T \mathbf{a}_i \tag{28}$$

Gram-Schmidt Orthonormalization (contd.)

- This iterative procedure gives us the orthogonal vectors $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n$.
- Finally, the orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ are obtained by the simple expression

$$\mathbf{q}_i = \frac{1}{||\mathbf{t}_i||}\mathbf{t}_i \tag{29}$$

- Let A be the matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ and Q, the matrix with columns $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$.
- It can be proved that $C(V) = C(Q)$, that is, the matrices V and Q have the same column space. The vector a_i can be expressed as

$$\mathbf{a}_i = \sum_{k=1}^{n} (\mathbf{a}_i^T \mathbf{q}_k)\mathbf{q}_k \tag{30}$$

- The i^{th} column of A is a linear combination of the columns of Q, with the scalar coefficient $\mathbf{a}_i^T \mathbf{q}_k$ for the k^{th} column of Q.

Gram-Schmidt Orthonormalization (contd.)

- By the very construction procedure of the Gram-Schmidt orthonormalization process, $\mathbf{a}_i$ is orthogonal to $\mathbf{q}_k$ for all $k > i$. Therefore, (30) can be expressed more precisely as

$$\mathbf{a}_i = \sum_{k=1}^{i} (\mathbf{a}_i^T \mathbf{q}_k)\mathbf{q}_k \tag{31}$$

- Therefore, matrix A can be decomposed into the product of Q with a upper triangular matrix R; $A = QR$, with $R_{k,i} = \mathbf{a}_i^T \mathbf{q}_k$. Since $\mathbf{a}_i^T \mathbf{q}_k = 0,\ \forall\ k > i$, we can easily see that R is upper traingular.

End Recap: Basis and Dimensions from Linear Algebra wrt $< .,. >_E$ (Eucledian Inner Product) (For your homework)

HW: In $\Re^n$, why $\|\mathbf{u}\|_p$ may not have an inner product for $p \neq 2$?

Proof:

- And here is how you can create a basis for $\mathcal{V}, < .,. >$:
  - If $\mathbf{u} = 0$ or $\mathbf{v} = 0$, then $< \mathbf{u}, \mathbf{v} >= 0$.
  - Assume $\mathbf{u}, \mathbf{v} \neq 0$. Let $\mathbf{z} = \mathbf{u} - \frac{<\mathbf{u},\mathbf{v}>}{<\mathbf{v},\mathbf{v}>}\mathbf{v}$.
  - By linearity of inner product in first argument, we have: $< \mathbf{z}, \mathbf{v} >=< \mathbf{u} - \frac{<\mathbf{u},\mathbf{v}>}{<\mathbf{v},\mathbf{v}>}\mathbf{v}, \mathbf{v} >=< \mathbf{u}, \mathbf{v} > - \frac{<\mathbf{u},\mathbf{v}>}{<\mathbf{v},\mathbf{v}>} < \mathbf{v}, \mathbf{v} >= 0$
  - Therefore, $< \mathbf{u}, \mathbf{u} >=< \mathbf{z} + \frac{<\mathbf{u},\mathbf{v}>}{<\mathbf{v},\mathbf{v}>}\mathbf{v}, \mathbf{z} + \frac{<\mathbf{u},\mathbf{v}>}{<\mathbf{v},\mathbf{v}>}\mathbf{v} >=< \mathbf{z}, \mathbf{z} > + (\frac{<\mathbf{u},\mathbf{v}>}{<\mathbf{v},\mathbf{v}>})^2 < \mathbf{v}, \mathbf{v} > +0$
  - So $< \mathbf{u}, \mathbf{u} >\geq \frac{|<\mathbf{u},\mathbf{v}>|^2}{<\mathbf{v},\mathbf{v}>}$

Compact representation of Inner Product Space

- Let the linear subspace $S \subseteq V$ be associated with an inner product $< ., . >$
- Let $B = basis(S)$ with respect to the arbitrary inner product $< ., . >$ (extending results from the eucledian inner product)
- Let $dim(V) = n$, and $dim(S) = m \leq n$.
- Define $S^{\perp}$; the orthogonal complement ($S^{\perp} \in V$) of S as:
  $S^{\perp} = \{v \in V \mid < v, u >= 0 \ \forall \ u \in S \ \}$
  This implies:-
  - Both S and $S^{\perp}$ are linear subspaces of V.
  - $S \cap S^{\perp} = \{0\}$, $dim(S) + dim(S^{\perp}) = n$
  - $(S^{\perp})^{\perp} = S$.
  - If $B^{\perp}$ is the basis for $S^{\perp}$, then $B \cup B^{\perp}$ is the basis for V.
  - $S = \{\mathbf{v} \in V | < \mathbf{v}, \mathbf{u} >= 0, \forall \ \mathbf{u} \in B^{\perp} \ \}$
  - $S^{\perp} = \{\mathbf{v} \in V \mid < \mathbf{v}, \mathbf{u} >= 0 \ \forall \ \mathbf{u} \in B \ \}$

HW: In $\Re^n$, why $\|\mathbf{u}\|_p$ may not have an inner product for $p \neq 2$?

Motivation:

- Consider the following inner product on $\Re^2$: For any $\mathbf{x}, \mathbf{y} \in \Re^2$, let $< \mathbf{x}, \mathbf{y} >= 2x_1y_1 - x_1y_2 - x_2y_1 + 4x_2y_2$. It can be easily verified that this in an inner product (by checking for linearity, symmetry and positive definiteness by expressing it as a sum of squares).
- This inner product is certainly different from the conventional (Eucledian) dot product $< \mathbf{x}, \mathbf{y} >_E = x_1y_1 + x_2y_2$ which corredponds to the $\|.\|_2$ norm.
- Is it possible that the $< \mathbf{x}, \mathbf{y} >$ defined in step 1 (or some other such inner product) corresponds to $\|.\|_p$ norm for $p \neq 2$?

HW: In $\Re^n$, why $\|\mathbf{u}\|_p$ may not have an inner product for $p \neq 2$?

Motivation:

- Consider the following inner product on $\Re^2$: For any $\mathbf{x}, \mathbf{y} \in \Re^2$, let $<\mathbf{x}, \mathbf{y}>= 2x_1y_1 - x_1y_2 - x_2y_1 + 4x_2y_2$. It can be easily verified that this in an inner product (by checking for linearity, symmetry and positive definiteness by expressing it as a sum of squares).
- This inner product is certainly different from the conventional (Eucledian) dot product $<\mathbf{x}, \mathbf{y}>_E = x_1y_1 + x_2y_2$ which corredponds to the $\|.\|_2$ norm.
- Is it possible that the $<\mathbf{x}, \mathbf{y}>$ defined in step 1 (or some other such inner product) corresponds to $\|.\|_p$ norm for $p \neq 2$?

In $\Re^n$, it can be proved that for any inner product vector space $(\mathcal{V}, <.,.>)$, the inner product $<.,.>$ (including the Eucledian one) can be represented as

$$<\mathbf{u}, \mathbf{v}>= \sum_{i=1}^{n}\sum_{j=1}^{n} a_ib_j <\mathbf{e}_i, \mathbf{e}_j>= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}^TE\mathbf{b} =<\mathbf{a}^T, \mathbf{b}>_E$$

HW: In $\Re^n$, why $\|\mathbf{u}\|_p$ may not have an inner product for $p \neq 2$?

Proof:

- In $\Re^n$, it can be proved that for any inner product vector space $(\mathcal{V}, < .,. >)$, the inner product $< .,. >$ (including the Eucledian one) can be represented as

$$< \mathbf{u}, \mathbf{v} >= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j < \mathbf{e}_i, \mathbf{e}_j >= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}^T E \mathbf{b} =< \mathbf{a}^T, \mathbf{b} >_E$$

  - Here, $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ is a basis for the inner product vector space.
  - The inner product $< .,. >_E$ is the eucledian inner product. That is, $< .,. >_E = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j$.

    The (positive definite) matrix E is defined as

$$E = \begin{bmatrix} < \mathbf{e}_1, \mathbf{e}_1 > & < \mathbf{e}_1, \mathbf{e}_2 > & \ldots\ldots & < \mathbf{e}_1, \mathbf{e}_n > \\ . & & & \\ < \mathbf{e}_n, \mathbf{e}_1 > & < \mathbf{e}_n, \mathbf{e}_2 > & \ldots\ldots & < \mathbf{e}_n, \mathbf{e}_n > \end{bmatrix} \quad (32)$$

  - Note that in any $\Re^n$, any inner product vector space $(\mathcal{V}, < .,. >)$ will have a basis of size at most n.

HW: In $\Re^n$, why $\|\mathbf{u}\|_p$ may not have an inner product for $p \neq 2$?

Thus, any inner product $< .,. >$ in $\Re^n$ can be expressed as a Eucledian inner product $< .,. >_E$, with possible rotation using a matrix R where $E = RR^T$ is a positive definite matrix[5]

[5]Recall from slides 25 to 27 that $\mathbf{x}^P\mathbf{x}$ is a norm if P is positive definite

Convex Sets

Motivations for Topology..(metric..norm...inner prod) => Neighborhod, Continuity

Motivation for inner product => Dual description...

Motivation for convex sets => You need convexity of the domain on which the convexity of the function is defined...

Convex sets

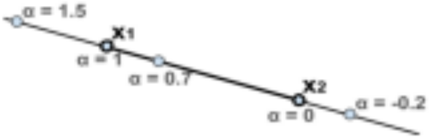
- affine and convex sets.
- some important examples
- operations that preserve convexity
- generalized inequalities
- separating and supporting hyperplanes
- dual cones and generalized inequalities

Affine set Can be thought of as a vector space shifted from the origin

- **In 2D, a line** through $\mathbf{x}_1, \mathbf{x}_2$: all points
-

$$\mathbf{x} = \alpha\mathbf{x}_1 + \beta\mathbf{x}_2$$

where $\alpha + \beta = 1, \alpha \in R.(1)$ $\alpha >= 0, \beta >= 0(2)$ $\alpha + \beta = 1$ and$\alpha >= 0, \beta >= 0(3)$



-
- **affine set** contains the line through any two distinct points in the set.
- **example** solution set of linear equations $\{\mathbf{x}|A\mathbf{x} = b\}$ This is an insight from linear algebra on geometry
  - No Solution: $\mathbf{x} = \phi$. Is that affine?
  - Unique Solution: $\mathbf{x}$ is a point.
  - Infinitely Many Solutions: $\mathbf{x}$ is a line, or a plane, etc.

(conversely every affine set can be expressed as solution set of system of linear equations)

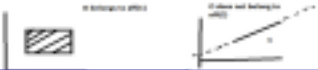
Convex set

- **In 2D, a line segment** between $\mathbf{x}_1$,$\mathbf{x}_2$: all points

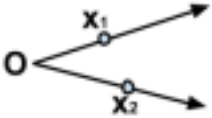
$$\mathbf{x} = \alpha \mathbf{x}_1 + \beta \mathbf{x}_2$$

$$\text{where} \quad \alpha + \beta = 1, 0 \leq \alpha \leq 1 (also, 0 \leq \beta \leq 1).$$

  (the term 'line segment' is more appropriate a name when $\mathbf{x}_1$ and $\mathbf{x}_2$ are points in real, finite dimensional Eucledian vector space $\Re^n$ or $\Re^{m \times n}$)
- **convex set** contains line segment between any two points in the set.
  - $\mathbf{x}_1, \mathbf{x}_2 \in C, 0 \leq \alpha \leq 1 \Rightarrow \alpha \mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2 \in C$
- **examples** (one convex, two non-convex sets).
  - 



Recap: Cone and conic combination



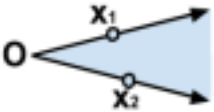
- **cone** A set C is a cone if $\forall \mathbf{x} \in C$, $\theta\mathbf{x} \in C$ for $\theta \geq 0$.
- **conic (nonnegative) combination** of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ is any point $\mathbf{x}$ of the form

$$\mathbf{x} = \theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 + ... + \theta_k\mathbf{x}_k$$

$$\text{with} \quad \theta_i \geq 0.$$

example : diagonal vector of a parallelogram is a conic combination of two vectors(points) $\mathbf{x}_1$ and $\mathbf{x}_2$ forming the parallelogram.

Conic hull(or convex cone) and Affine hull

- **Conic hull or conic(S):** The set that contains all conic combinations of points in set S.
- conic(S) = Smallest conic set that contains S.
- 

- Similarly, **Affine hull or aff(S):** The set that contains all affine combinations of points in set S.
- aff(S) = Smallest affine set that contains S.
-