Lecture 4:
Inner Product Theorems
Gram-Schmidt Orthogonalization
Subspaces

Date Revised: 2008/10/06
Date Given: 2008/10/06
Inner Product Theorems

The definition of inner product immediately gives rise to some useful results. We will only give the basic idea of the proofs; you can find the details in Shankar.

“Law of Cosines”

\[ |v + w|^2 = |v|^2 + |w|^2 + 2 \Re(\langle v | w \rangle) \]  

(3.19)

where \( \Re(z) \) is the real part of the complex number \( z \). This is proven by simply expanding out the inner product implied by the left side. The relation is named as it is because it reduces to the law of cosines when \( |v \rangle \) and \( |w \rangle \) belong to \( \mathbb{R}^3 \).

The astute reader will see that the sign on the inner product term is different than what one usually sees in the law of cosines, \( c^2 = a^2 + b^2 - 2ab \cos \gamma \). This is because the angle \( \gamma \) is not the angle between \( \vec{a} \) and \( \vec{b} \) at the origin; it is the supplementary angle to that angle, hence the sign flip on the cos term. The following diagram this explicitly.
Diagram to illustrate sign of last term in law of cosines in $\mathbb{R}^2$. The law of cosines conventionally involves the angle $\gamma$ because that is the angle opposite to the vector $\vec{a} + \vec{b}$. The inner product of two vectors gives the cosine of the angle between them when they are placed at the origin, $\theta$. There is a sign flip between the two cosines because they are supplementary angles (sum to $\pi$ radians). We want to use the $\theta$ angle instead of the $\gamma$ angle in the generic form of the law of cosines, hence the sign flip.
Schwarz Inequality

\[ |\langle v | w \rangle|^2 \leq |v|^2 |w|^2 \]  

(3.20)

The Schwarz inequality simply states that the inner product of two vectors can be no larger than the product of their lengths. More simply, if one divides out the norms of the vectors, it states that the inner product of two unit vectors can be no greater in magnitude than 1. When we think about the interpretation of the inner product as the projection of one vector onto the other, this makes sense; the projection of a unit vector onto another unit vector can be no larger than 1.

Diagram illustrating Schwarz inequality in \( \mathbb{R}^2 \) for vectors of unit length. All three vectors \( \vec{a}, \vec{b}_1, \) and \( \vec{b}_2 \) have unit length (as indicated by the fact that they all end on the circle). Clearly, the projections \( |\langle \vec{a} | \vec{b}_1 \rangle| \) and \( |\langle \vec{a} | \vec{b}_2 \rangle| \) are both less than 1. Note how the sign of the projection does not affect the result.
The inequality is proven by applying the law of cosines to the vector

\[ |v'\rangle = |v\rangle - \langle w | v \rangle \frac{|w\rangle}{|w|^2} = |v\rangle - \langle \hat{w} | v \rangle |\hat{w}\rangle \]

where \( |\hat{w}\rangle = \hat{w}/|w| \) is the unit vector along the \( |w\rangle \) direction, and using the positive definiteness of the norm of \( |v'\rangle \).

Clearly, \( |v'\rangle \) is the piece of \( |v\rangle \) that is orthogonal to \( |w\rangle \). The Schwarz inequality devolves to an equality if \( |v\rangle = \lambda |w\rangle \) for some \( \lambda \); i.e., if \( |v\rangle \) and \( |w\rangle \) are the same up to a multiplicative constant, indicating they point in the same (or opposite) direction.

Note that the above vector may also be written as

\[ |v'\rangle = |v\rangle - |\hat{w}\rangle\langle \hat{w} | v \rangle \]

We see the expression \( |\hat{w}\rangle\langle \hat{w} | \) as we did when writing out bras and kets as sums of components along the vectors of an orthonormal basis. Such objects we will see are projection operators because they project out the part of the vector they operate on along the unit vector comprising the operator.
Triangle Inequality

\[ |v + w| \leq |v| + |w| \]  \hspace{1cm} (3.23)

This is a direct result of the law of cosines, arising from the fact that
\[ 2 \mathcal{R} (\langle v | w \rangle) \leq 2 |\langle v | w \rangle| \leq 2 |v||w|. \] The inequality devolves to an equality only if
\[ |v\rangle = \lambda |w\rangle \] with \( \lambda \) real and positive.

Diagram illustrating triangle inequality in \( \mathbb{R}^2 \), where it expresses the fact that the length of the sum of two vectors can be no larger than the sum of their individual lengths, and equality occurs when the vectors are coaligned. The circle is centered on the start of \( \vec{b} \) and has radius equal to \( |\vec{b}| \), so indicates the locus of possible endpoints of \( \vec{a} + \vec{b} \); one particular example is given for the orientation of \( \vec{b} \) and \( \vec{a} + \vec{b} \). The dashed line indicates the maximum length possibility, with \( \vec{a} \) and \( \vec{b} \) coaligned so that
\[ |\vec{a} + \vec{b}| = |\vec{a}| + |\vec{b}|. \]
Gram-Schmidt Orthogonalization

Given \( n \) linearly independent vectors \( \{|v_j\rangle\} \), one can construct from them an orthonormal set of the same size.

The basic idea is to orthogonalize the set by subtracting from the \( j \)th vector the projections of that vector onto the \( j-1 \) previous vectors, which have already been orthogonalized. Dirac notation makes the projection operations more obvious.

We begin by using the first ket from the original set, creating a normalized version:

\[
|1\rangle = |v_1\rangle \\
|1\rangle = \frac{|1\rangle}{\sqrt{\langle1'|1\rangle}}
\] (3.24)

Then, the second member of the orthogonal and orthonormal sets are

\[
|2\rangle = |v_2\rangle - \frac{|1\rangle\langle1'|v_2\rangle}{\langle1'|1\rangle} = |v_2\rangle - |1\rangle\langle1|v_2\rangle \\
|2\rangle = \frac{|2\rangle}{\sqrt{\langle2'|2\rangle}}
\] (3.25)

and so on; the generic formula is

\[
|j\rangle = |v_j\rangle - \sum_{k=1}^{j-1} \frac{|k\rangle\langle k'|v_k\rangle}{\langle k'|k\rangle} = |v_j\rangle - \sum_{k=1}^{j-1} |k\rangle\langle k|v_k\rangle \\
|j\rangle = \frac{|j\rangle}{\sqrt{\langle j'|j\rangle}}
\] (3.26)
Diagram illustrating Gram-Schmidt orthogonalization in $\mathbb{R}^2$. The first vector $|v_1\rangle$ is simply normalized to obtain $|1\rangle$ but otherwise left unchanged. We subtract off from the second vector $|v_2\rangle$ the projection along $|1\rangle$, leaving $|2'\rangle$. We then normalize $|2'\rangle$ to obtain $|2\rangle$. $|1\rangle$ and $|2\rangle$ are clearly orthogonal and normalized. (The circle has unity radius.)

One proves this theorem inductively, showing that if the first $j$-1 vectors have been orthogonalized, then the $j$th vector created via the above formula is orthogonal to the first $j$-1. The $|j\rangle$ are manifestly normalized.

Gram-Schmidt orthogonalization lets us conclude what we intuitively expect: for an inner product space of dimension $n$, we can construct an orthonormal basis for the space from any other basis. Shankar proves this point rigorously, but it is easy to see intuitively: the Gram-Schmidt procedure tells us that any linearly independent set of $n$ vectors yields a mutually orthogonal set of $n$ vectors, and it is fairly obvious that a mutually orthogonal set of $n$ vectors is linearly independent. If the initial set of LI vectors matches the dimension of the space, then the new orthonormal set is a basis for the space.
Subspaces

It almost goes without saying that a subspace of a linear vector space or inner product space is a subset of the space that is itself a vector or inner product space. Since the subset inherits the algebraic operations (and inner product, if one exists) from the parent space, the only substantive requirement is that the subspace be closed under the vector addition and scalar-vector multiplication operations. One can show that the parent space's null vector must be in any subspace.

Given two subspaces $V_1$ and $V_2$ of a vector space $V$, the sum or direct sum of the two subspaces, denoted by $V_1 \oplus V_2$, is the set of all linear combinations of vectors in $V_1$ and $V_2$. Note that, since $V_1$ and $V_2$ are subspaces of some larger vector space $V$, it is already known that one may add vectors from $V_1$ and $V_2$ together.

Note that $V_1 \oplus V_2$ is not the same as $V_1 \cup V_2$. $V_1 \oplus V_2$ consists of all linear combinations of the form

$$\left| \nu \right\rangle = \alpha_1 \left| v_1 \right\rangle + \alpha_2 \left| v_2 \right\rangle$$

(3.27)

where $\left| v_1 \right\rangle$ is in $V_1$ and $\left| v_2 \right\rangle$ is in $V_2$. When both $\alpha_1$ and $\alpha_2$ are nonzero, $\left| \nu \right\rangle$ belongs to neither $V_1$ nor $V_2$, but lives in the part of $V_1 \oplus V_2$ outside $V_1$ and $V_2$. On the other hand, $V_1 \cup V_2$ consists only of the linear combinations for which at least one of $\alpha_1$ and $\alpha_2$ vanish (i.e., the trivial linear combination in which no combining is done!).

The following diagram may help to illustrate the distinctions.
Diagram illustrating subspaces in $\mathbb{R}^2$. The first subspace $V_1$ consists of all vectors along the $x$-axis only, indicated by the red line. The second subspace $V_2$ consists of vectors along the $y$-axis only, indicated by the blue line. The union $V_1 \cup V_2$ consists of all vectors either along the red line or the blue line. The direct sum $V_1 \oplus V_2$ consists of all linear combinations of vectors along the red line or the blue line, so covers the entire plane, indicated by the pink shading. $V_1 \oplus V_2$ is much bigger than $V_1$, $V_2$, or $V_1 \cup V_2$.

The most trivial subspace an inner product space $V$ can have is the set of all vectors of the form $\alpha |v\rangle$ where $|v\rangle$ is some element in $V$: these are just all the vectors along $|v\rangle$. Given a basis $\{|v_j\rangle\}$, the entire space $V$ is just the direct sum of the subspaces of this type for each basis element. That is, if we define $V_j = \{\alpha |v_j\rangle\}$, the set of all scalar multiples of the $j$th basis element, then

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_N$$
Example 3.16: $\mathbb{R}^N$ and $\mathbb{C}^N$

Each of these have a variety of subspaces. Each can be viewed as being the $N$-element direct sum of their $N = 1$ versions (à la what we just said):

$$\mathbb{R}^N = \mathbb{R} \oplus \mathbb{R} \cdots \oplus \mathbb{R}$$

$$\mathbb{C}^N = \mathbb{C} \oplus \mathbb{C} \cdots \oplus \mathbb{C}$$

or, perhaps a direct sum of many copies of $\mathbb{R}^2$ with one $\mathbb{R}$ thrown in if $N$ is odd, or direct sums of $\mathbb{R}^M$ of various $M$, etc., etc.

Example 3.17: The set of all complex-valued functions on a set of discrete points $i \frac{L}{n+1}, i = 1, \ldots, n$, in the interval $(0, L)$, as in Example 3.4

The analogue of the above when considering the function representation would be to select functions that are zero at various points. For example, if $N = 4$, the functions that are always zero on $x_1$ and $x_2$ are one subspace, the functions that always vanish on $x_3$ and $x_4$ are a different subspace, and the full space is the direct sum of these two subspaces. Each subspace is isomorphic to a function space on two discrete points, and hence to $\mathbb{C}^2$. 
Example 3.18: Spin-1/2 particle at the origin

This is just $\mathbb{C}^2$ and so the subspaces are fairly boring, but the physical interpretation is interesting. One can consider the two subspaces \{\alpha | \uparrow_z \rangle\} and \{\alpha | \downarrow_z \rangle\} where $\alpha$ is any complex number: these subspaces consist of either spin up or spin down states only. (Note that, even if you restrict to $|\alpha| = 1$, there are still an infinity of elements in each subspace because $\alpha$ is complex.) But one could alternately consider the subspaces \{\alpha | \uparrow_x \rangle\} and \{\alpha | \downarrow_x \rangle\} or \{\alpha | \uparrow_y \rangle\} and \{\alpha | \downarrow_y \rangle\}. One recovers the full space by direct sum of the two subspaces in each circumstance, but these provide some alternate subspaces of $\mathbb{C}^2$. 
Example 3.19: Real antisymmetric, imaginary symmetric, and anti-Hermitian matrices.

The aforementioned set of real, antisymmetric $N \times N$ matrices with a real number field form a subspace of the set of anti-Hermitian matrices\(^a\), also with a real field. The set of purely imaginary $N \times N$ symmetric matrices with a real field are also a subspace of the anti-Hermitian matrices. The direct sum of the real antisymmetric matrices and the purely imaginary symmetric matrices gives the entire space of anti-Hermitian matrices. Specifically, the three groups are

$$A_R = \begin{bmatrix} 0 & a_{12} & a_{13} \\ -a_{12} & 0 & a_{23} \\ -a_{13} & -a_{23} & 0 \end{bmatrix} \quad A_I = \begin{bmatrix} 0 & i b_{12} & i b_{13} \\ i b_{12} & 0 & i b_{23} \\ i b_{13} & i b_{23} & 0 \end{bmatrix} \quad A_A = \begin{bmatrix} 0 & a_{12} + i b_{12} & a_{13} + i b_{13} \\ -a_{12} + i b_{12} & 0 & a_{23} + i b_{23} \\ -a_{13} + i b_{13} & -a_{23} + i b_{23} & 0 \end{bmatrix}$$

\(^a\)Complex matrices $A$ for which $(A^*)^T = -A$ where * is element-by-element complex conjugation and $^T$ is matrix transposition.
Let’s check a number of things:

▶ Real antisymmetric matrices are a subset of anti-Hermitian matrices because they do not change under complex conjugation and they pick up a sign under transposition. Similarly, purely imaginary symmetric matrices are a subset because they pick up a sign under complex conjugation but do not change under transposition.

▶ We have already shown that real antisymmetric matrices are closed. Purely imaginary matrices symmetric matrices are closed under addition and multiplication by real numbers because neither operation can change the fact they are purely imaginary or symmetric.

▶ We have already shown that real antisymmetric matrices are an inner product space. Purely imaginary symmetric matrices are also an inner product space because the complex conjugation in the inner-product formula ensures positive definiteness. The transpose and linearity rules are also satisfied.

▶ Any sum of a real antisymmetric matrix and a purely imaginary symmetric matrix is immediately anti-Hermitian because the real part of the sum is guaranteed to change sign and the imaginary part to keep its sign under transposition. Any anti-Hermitian matrix can be decomposed in terms of a real antisymmetric and purely imaginary symmetric matrix simply by breaking it element-by-element into real and imaginary parts.