In the previous note we learned an important algorithm to produce orthogonal sequences of vectors called the Gramm-Schmidt orthogonalization process.

Gramm-Schmidt orthogonalization process. Let $< \, , >$ be an inner product on a vector space V and let $v_1, \ldots, v_k$ be linearly independent. Define recursively vectors $w_1, \ldots, w_k$ as follows:

$$w_1 = v_1, \quad w_{j+1} = v_{j+1} - \left(\frac{< v_{j+1}, w_1 >}{< w_1, w_1 >} w_1 + \frac{< v_{j+1}, w_2 >}{< w_2, w_2 >} w_2 + \ldots + \frac{< v_{j+1}, w_j >}{< w_j, w_j >} w_j \right).$$

Then

1. $w_1, w_2, \ldots, w_k$ is orthogonal.
2. $\mathrm{span}\{w_1, \ldots, w_j\} = \mathrm{span}\{v_1, \ldots, v_j\}$ for $j = 1, \ldots, k$.

As we noted before, when the process is applied to linearly dependent vectors $v_1, \ldots, v_k$ then we will get $w_j = 0$ for some j (in fact, j is the smallest index such that v_j is a linear combination of $v_1, \ldots, v_{j-1}$) and then the process can not be continued.

Another simple observation is that if the first j vectors of the sequence $v_1, \ldots v_k$ are already orthogonal to each other (i.e. the sequence $v_1, \ldots v_j$ is orthogonal) then the Gramm-Schmidt process will yield $w_1 = v_1$, $w_2 = v_2$,..., $w_j = v_j$.

The Gramm-Schmidt orthogonalization has several important consequences which we are going to discuss now. We will use the term **inner product space** for a vector space with an inner product.

Theorem. Every finite dimensional inner product space has an orthonormal basis.

Indeed, start with an arbitrary basis $v_1, \ldots, v_n$ and apply the Gramm-Schmidt orthogonalization process to get an orthogonal sequence $w_1, \ldots, w_n$. We know that any orthogonal sequence is linearly independent so $w_1, \ldots, w_n$ is an orthogonal basis (since $n = \dim V$). Then $\frac{w_1}{||w_1||}, \frac{w_2}{||w_2||}, \ldots, \frac{w_n}{||w_n||}$ is an orthonormal basis.

Theorem. Let V be a finite dimensional inner product space. Every orthogonal (orthonormal) sequence in V can be extended to an orthogonal (orthonormal) basis.

Indeed, suppose that $w_1, \ldots, w_k$ is orthogonal. Then it is linearly independent hence can be extended to a basis $w_1, \ldots, w_k, v_{k+1}, \ldots, v_n$. The Gramm-Schmidt process applied to this basis yields an orthogonal basis $w_1, \ldots, w_k, w_{k+1}, \ldots, w_n$ extending our original sequence (we know that the first k vectors will be unchanged by the Gramm-Schmidt process). If we started with an orthonormal sequence $w_1, \ldots, w_k$ then $w_1, \ldots, w_k, \frac{w_{k+1}}{||w_{k+1}||}, \ldots, \frac{w_n}{||w_n||}$ is an orthonormal basis extending our sequence.

Orthogonal bases are very convenient for computations:

Proposition. Let $v_1, \ldots, v_n$ be an orthogonal basis of an inner product space V.

1. For any vector $v \in V$ we have
$$v = \frac{< v_1, v >}{< v_1, v_1 >} v_1 + \frac{< v_2, v >}{< v_2, v_2 >} v_2 + \ldots + \frac{< v_n, v >}{< v_n, v_n >} v_n.$$
2. If $T : V \longrightarrow V$ is linear transformation then the matrix $A = [a_{i,j}]$ representing T in the basis $v_1, \ldots, v_n$ has entries $a_{i,j} = \dfrac{< v_i, T(v_j) >}{< v_i, v_i >}$.

Indeed, $v = a_1 v_1 + \ldots + a_n v_n$ for some scalars $a_1, \ldots, a_n$. Thus
$$< v_i, v >=< v_i, a_1 v_1 + \ldots + a_n v_n >= a_1 < v_i, v_1 > + \ldots + a_n < v_i, v_n >= a_i < v_i, v_i >$$
since $< v_i, v_j >= 0$ for $j \neq i$. It follows that $a_i = \frac{<v_i,v>}{<v_i,v_i>}$ for $i = 1, \ldots, n$ which proves part 1.

For part 2 recall that $a_{i,j}$ is the $i-th$ coordinate of $T(v_j)$ in the basis $v_1, \ldots, v_n$. Part 1 tells us that the i-th coordinate of any vector v in the basis $v_1, \ldots, v_n$ is equal to $\frac{<v_i,v>}{<v_i,v_i>}$. Thus $a_{i,j} = \frac{<v_i,T(v_j)>}{<v_i,v_i>}$.

The above formulas become even simpler if the basis is orthonormal, as then $< v_i, v_i >= 1$:

Proposition. Let $v_1, \ldots, v_n$ be an orthonormal basis of an inner product space V.

1. For any vector $v \in V$ we have
$$v =< v_1, v > v_1 + < v_2, v > v_2 + \ldots + < v_n, v > v_n.$$
2. If $T : V \longrightarrow V$ is linear transformation then the matrix $A = [a_{i,j}]$ representing T in the basis $v_1, \ldots, v_n$ has entries $a_{i,j} =< v_i, T(v_j) >$.
3. if $u, w \in V$ have in the basis $v_1, \ldots, v_n$ coordinates $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ respectively, then $< u, w >= a_1 b_1 + \ldots + a_n b_n$. In other words, the inner product expressed in the coordinates with respect to an orthonormal basis is just the dot product of the vectors of coordinates.

Indeed, parts 1 and 2 are just special cases of the previous proposition. Part 3 is a simple computation. First note that $u = a_1 v_1 + \ldots + a_n v_n$ and $w = b_1 v_1 + \ldots + b_n v_n$. Thus $< u, w >= a_1 < v_1, w > + \ldots + a_n < v_n, w >$. By part 1, $< v_i, w >= b_i$ for $i = 1, \ldots, n$, so the result follows.

Note that part 3 tells us that any two inner product spaces of the same dimension look the same when expressed in an orthonormal basis. A more precise way to say this is the following result.

Theorem. Let V and W inner product spaces of the same finite dimension. Let $v_1, \ldots v_n$ be an orthonormal basis of V and let $w_1, \ldots, w_n$ be an orthonormal basis of W. The linear transformation $T : V \longrightarrow W$ defined by the condition $T(v_i) = w_i$ for $i = 1, 2, \ldots, n$ is an isomorphism which preserves inner product, i.e.
$$< T(u), T(v) >_W =< u, v >_V \quad \text{for any } u, v \in V$$
(here $< \, , \, >_V, < \, , \, >_W$ denote the inner product on V and W respectively).

Indeed, coordinates of any vector $v \in V$ in the basis $v_1, \ldots v_n$ are exactly the same as the coordinates of $T(v)$ in the basis $w_1, \ldots, w_n$ and the inner products are equal to the dot product of the vectors of coordinates.

This result says that any two inner product spaces of the same dimension can be identified (are isomorphic). However there are many such identifications (since there are many choices for an orthonormal basis), none of which can be chosen as preferred in any meaningful way. This is why we study the more abstract concept of an inner product space rather than just focus on $\mathbb{R}^n$ with the dot product.

The orthogonality relation has many nice properties. To state them we introduce the following definition

Definition. Let V be an inner product space and S a subset of V. We define the **orthogonal complement** $S^\perp$ of S as the set of all vectors in V which are orthogonal to every vector in S:

$$S^\perp = \{v \in V : v \perp w \text{ for all } w \in S\}.$$

This concept has the following properties

Theorem. Let V be a finite dimensional inner product space.

1. $S^\perp$ is a subspace of V for any subsets S.
2. If $S \subseteq T$ then $T^\perp \subseteq S^\perp$.
3. $(S^\perp)^\perp = \operatorname{span}(S)$.
4. $(\operatorname{span}(S))^\perp = S^\perp$.
5. $S \cap S^\perp \subseteq \{0\}$.
6. If W is a subspace of V then any vector $v \in V$ can be written in a unique way as $w + u$, where $w \in W$ and $u \in W^\perp$.
7. If W is a subspace of V then $\dim W + \dim W^\perp = \dim V$.

Let us justify all these properties. Let S be a subset of V. Since 0 is perpendicular to all vectors, we have $0 \in S^\perp$. Suppose $v, w \in S^\perp$. Then for any $u \in S$ we have $< v+w, u >=< v, u > + < w, v >= 0+0 = 0$. It follows that $v + w$ is orthogonal to all vectors in S, i.e. $v + w \in S^\perp$. Similarly, for any scalar a we have $< av, u >= a < v, u >= a \cdot 0 = 0$, so $av \in S^\perp$. This proves that $S^\perp$ is a subspace of V.

Clearly if a vector is perpendicular to very vector in T then it is perpendicular to every vector in the subset S of T, This show that $T^\perp \subseteq S^\perp$.

By the very definition, every vector in S is orthogonal to any vector in $S^\perp$. It follows that $S \subseteq (S^\perp)^\perp$. Since $(S^\perp)^\perp$ is a subspace by part 1, we have $\operatorname{span}(S) \subseteq (S^\perp)^\perp$. Choose an orthonormal basis $v_1, \dots v_k$ of $\operatorname{span}(S)$ and complete it to an orthonormal basis of V by adding $v_{k+1} \dots, v_n$. For $j > k$ the subspace ${v_j}^\perp$ contains vectors $v_1, \dots, v_k$, hence it contains the whole space $\operatorname{span}\{v_1, \dots, v_k\} = \operatorname{span}(S)$. Thus, for $j > k$, the vector v_j is orthogonal to every vector in $\operatorname{span}(S)$, hence also to every vector in S. In other words, $v_j \in S^\perp$ for $j > k$. Suppose now that $v = \sum_{i=1}^n a_i v_i \in (S^\perp)^\perp$. Then $0 =< v, v_j >= a_j$ for $j > k$. Thus $v = \sum_{i=1}^k a_i v_i \in \operatorname{span}(S)$. This shows that $(S^\perp)^\perp \subseteq \operatorname{span}(S)$. Since we earlier showed the opposite inclusion, we have the equality claimed in property 3.

By property 3, we have $((S^\perp)^\perp)^\perp = \operatorname{span}(S^\perp) = S^\perp$ (since $S^\perp$ is a subspace). On the other hand, again by property 3, $((S^\perp)^\perp)^\perp = (\operatorname{span}(S))^\perp$. This shows property 4.

If a vector v belongs to both S and $S^\perp$ then $v \perp v$, which happens only for $v = 0$. This shows property 5.

Now we justify property 6. Suppose first that we have two ways of writing v as a sum of a vector from W and a vector from $W^\perp$: $v = w + u = w_1 + u_1$. Then $w - w_1 = u_1 - u$. But $w - w_1 \in W$ and $u_1 - u \in W^\perp$

so the vector $w - w_1 = u_1 - u$ is in $W \cap W^{\perp}$. By property 4 we conclude that this vector is 0 so $w = w_1$ and $u = u_1$. This shows that there is at most one way of expressing v as a sum of a vector from W and a vector from $W^{\perp}$. Now choose an orthonormal basis $w_1, \ldots, w_k$ of W and complete it to an orthonormal basis of V by adding $w_{k+1}, \ldots, w_n$. For $j > k$ the vector w_j is orthogonal to $w_1, \ldots, w_k$, hence it is in $W^{\perp}$. Now $v = a_1 w_1 + \ldots + a_n w_n$ for some scalars $a_1, \ldots, a_n$. Note that $w = a_1 w_1 + \ldots a_k w_k \in W$, $u = a_{k+1} w_{k+1} + \ldots a_n w_n \in W^{\perp}$, and $v = w + u$. This shows existence of w and u.

Note that if $v \in W^{\perp}$ then $v = 0 + v$ is the decomposition from property 6, i.e. $w = 0$, $u = v$. We showed above that u is a linear combination of $w_{k+1}, \ldots, w_n$. Thus any $v \in W^{\perp}$ is a linear combination of $w_{k+1}, \ldots, w_n$, so $w_{k+1}, \ldots, w_n$ is a basis of $W^{\perp}$. This shows that $\dim W^{\perp} = n - k = \dim V - \dim W$, i.e. we get 7.

Let us discuss now some consequences of property 6. First a definition.

Definition. Let V be an inner product space and W a subspace of V. For $v \in V$ define $\mathrm{P}_W(v)$ to be the unique vector $w \in W$ such that $v = w + u$ for some $u \in W^{\perp}$. Thus $\mathrm{P}_W : V \longrightarrow W$ is a function which is called **the orthogonal projection** onto W.

We will often think of P_W as a function $\mathrm{P}_W : V \longrightarrow V$ (whose image is in the subspace W).

Proposition. Let V be an inner product space and W a subspace of V.

1. $\mathrm{P}_W : V \longrightarrow V$ is a linear transformation.
2. $\ker \mathrm{P}_W = W^{\perp}$, image$(\mathrm{P}_W) = W$, and $\mathrm{P}_W \circ \mathrm{P}_W = \mathrm{P}_W$.
3. If $w_1, \ldots, w_k$ is an orthogonal basis of W then

$$\mathrm{P}_W(v) = \frac{< v, w_1 >}{< w_1, w_1 >} w_1 + \frac{< v, w_2 >}{< w_2, w_2 >} w_2 + \ldots + \frac{< v, w_k >}{< w_k, w_k >} w_k.$$

Indeed, if $v_1 = z_1 + u_1$, $v_2 = z_2 + u_2$ with $z_1, z_2 \in W$ and $u_1, u_2 \in W^{\perp}$ then $v_1 + v_2 = (z_1 + z_2) + (u_1 + u_2)$ and $av_1 = az_1 + au_1$ for any scalar a. Note that $z_1 + z_2$ and az_1 are in W and $u_1 + u_2, au_2 \in W^{\perp}$. It follows that

$$\mathrm{P}_W(v_1 + v_2) = z_1 + z + 2 = \mathrm{P}_W(z_1) + \mathrm{P}_W(z_2), \text{ and } \mathrm{P}_W(av_1) = az_1 = a\mathrm{P}_W(v_1).$$

This shows that P_W is linear.

Clearly $\mathrm{P}_W(v) = 0$ if and only if $v = 0 + u$ for some $u \in W^{\perp}$, i.e. if and only if $v \in W^{\perp}$. This shows that $\ker \mathrm{P}_W = W^{\perp}$. By the very definition, the image of P_W is contained in W. On the other hand, if $w \in W$ then $w = w + 0$ and 0 is in $W^{\perp}$, so $\mathrm{P}_W(w) = w$. This means that all vectors in W are in the image of P_W, hence image$(\mathrm{P}_W) = W$. We have seen that $\mathrm{P}_W(w) = w$ for $w \in W$. Since $\mathrm{P}_W(v)$ is in W for any $v \in V$, we have $\mathrm{P}_W(\mathrm{P}_W(v)) = \mathrm{P}_W(v)$. Thus $\mathrm{P}_W \circ \mathrm{P}_W = \mathrm{P}_W$.

If $v \in W$ then we know that $v = \frac{<v,w_1>}{<w_1,w_1>} w_1 + \frac{<v,w_2>}{<w_2,w_2>} w_2 + \ldots + \frac{<v,w_k>}{<w_k,w_k>} w_k$ and $v = \mathrm{P}_W(v)$, so property 3 holds in this case. If $v \notin W$ then $w_1, \ldots, w_k, v$ is linearly independent. The Gramm-Schmidt process applied to this sequence will yield $w_1, \ldots, w_k, u$ where $u \in W^{\perp}$. The formula for u yields $v = \frac{<v,w_1>}{<w_1,w_1>} w_1 + \frac{<v,w_2>}{<w_2,w_2>} w_2 + \ldots + \frac{<v,w_k>}{<w_k,w_k>} w_k + u$. Since the vector $\frac{<v,w_1>}{<w_1,w_1>} w_1 + \frac{<v,w_2>}{<w_2,w_2>} w_2 + \ldots + \frac{<v,w_k>}{<w_k,w_k>} w_k$ is in W, it is equal to $\mathrm{P}_W(v)$.

Property 3 provides a more geometric view on the Gramm-Schmidt orthogonalization process: the vector w_k is equal to $v_k - u$, where u is the projection of v onto the subspace span$\{v_1, \ldots, v_{k-1}\}$.

The orthogonal projection of a vector v onto a subspace W has an important feature: it minimizes the length $||v - w||$ among all $w \in W$. To see this we first state a simple but important result, which extends a classical result from Euclidean geometry.

Pythagoras Theorem. Vectors v and w are orthogonal if and only if $||v+w||^2 = ||v||^2 + ||w||^2$.

Indeed,

$$||v+w||^2 =< v+w, v+w >=< v, v > + < w, w > +2 < v, w >= ||v||^2 + ||w||^2 + 2 < v, w >$$

so the equality $||v+w||^2 = ||v||^2 + ||w||^2$ holds if and only if $< v, w >= 0$.

Theorem. Let V be an inner product space, let W a subspace of V, and let $v \in V$. Among all vectors $w \in W$ the distance $||v-w||$ is shortest possible if and only if $w = \mathrm{P}_W(v)$.

Indeed, write $v = \mathrm{P}_W(v) + u$ for some $u \in W^\perp$. For any $w \in W$ the vectors $\mathrm{P}_W(v) - w$ and u are orthogonal. Thus, by Pythagoras, we have $||v-w||^2 = ||(\mathrm{P}_W(v)-w)+u||^2 = ||(\mathrm{P}_W(v)-w)||^2 + ||u||^2 \geq ||u||^2$ with equality if and only if $\mathrm{P}_W(v) - w = 0$.