When we work with abstract vector spaces, we usually chose a basis, and then represent vectors via their coordinates in the chosen basis. Then we compute with the coordinates as we do with vectors in $\mathbb{R}^n$. Since there is no preferred choice of a basis, we need to be able to compare computations in one basis to similar computations in another basis. This is done via the transition matrices which provide a "translation" from one basis to another. We can then look for a choice of a basis best suited for a particular problem we need to consider.

Similar philosophy can be applied to inner product spaces. In the last note we have seen that when working with inner product spaces it is natural and very convenient to use orthonormal bases. One can say that orthonormal basis for inner product spaces play the same role as arbitrary bases do for abstract vector spaces. Since there are many choices for an orthonormal basis,it is natural to study transition matrices from one orthonormal basis to another. It turns out that such matrices have very special properties.

Suppose that V is an inner product space with orthonormal basis $v_1, \ldots, v_n$ and also with another orthonormal basis $w_1, \ldots, w_n$. We will denote the first basis by B and the second by D. Recall that the i, j-entry of the transition matrix ${}_DI_B$ is the i-th coordinate of v_j in the basis D. In the last not we have seen that the i-th coordinate of any vector v in an orthonormal basis $w_1, \ldots, w_n$ is equal to $< w_i, v >$. Thus the i, j-entry of ${}_DI_B$ is equal to $< w_i, v_j >$.

Replacing the roles of B and D in the above discussion, we see that the i, j-entry of the matrix ${}_BI_D$ is equal to $< v_i, w_j >$, which is the same as $< w_j, v_i >$, which in turn is the j, i-entry of ${}_DI_B$. This shows that ${}_BI_D$ is the transpose of ${}_DI_B$. On the other hand, ${}_BI_D$ is the inverse of ${}_DI_B$. Thus the transpose of ${}_DI_B$ is the same as the inverse of ${}_DI_B$. This prompts the following definition.

Definition. A square matrix A is called **orthogonal** when the transpose of A is the inverse of A, i.e. if $AA^t = I$ (equivalently, if $A^{-1} = A^t$).

Thus the transition matrix from one orthonormal basis to another is an orthogonal matrix. In fact, we have the following result.

Proposition. Let $A = [a_{i,j}]$ be an $n \times n$ matrix. The following conditions are equivalent.

1. A is orthogonal.
2. A is a transition matrix from some orthogonal basis to another.
3. If $v_1, \ldots, v_n$ is an orthonormal basis of an inner product space then the vectors $w_j = a_{1,j}v_1 + a_{2,j}v_2 + \ldots + a_{n,j}v_n$, $j = 1, \ldots, n$ form another orthonormal basis and A is the transition matrix form the basis $w_1, \ldots, w_n$ to the basis $v_1, \ldots, v_n$.
4. The columns of A considered as vectors of in $\mathbb{R}^n$ form an orthonormal basis of $\mathbb{R}^n$ (with dot product as an inner product).
5. The rows of A considered as vectors of in $\mathbb{R}^n$ form an orthonormal basis of $\mathbb{R}^n$ (with dot product as an inner product).

Indeed, we already know that 2 implies 1.

Suppose that A is orthogonal. Then $AA^t = I$. The i, j-entry of AA^t is the dot product of of the i-th row of A and the j-th row of A (which is the same as the j-th column of A^t). Thus $r_i \cdot r_j = 0$ if $i \neq j$ and $r_i \cdot r_i = 1$ (here r_i stands for the i=th row of A). This proves 5. So 1 implies 5.

Suppose A satisfies 5. Then, as we have just seen, $r_i \cdot r_j = 0$ if $i \neq j$ and $r_i \cdot r_i = 1$. This means that $AA^t = I$. It follows that $A^tA = I$. But the i, j-entry of A^tA is the dot product of the i-th column of A and the j-th column of A. This means that 4 holds for A. Thus 5 implies 4.

Suppose A satisfies 4. Let v_i and w_i be as in property 3. Then the j-th column of A consists of coordinates of w_j in the basis $v_1, \ldots, v_n$. Since $v_1, \ldots, v_n$ is an orthonormal basis, the inner product $< w_i, w_j >$ is the same as the dot product of the coordinate vectors of w_i and w_j in the basis $v_1, \ldots, v_n$. In other words, $< w_i, w_j >$ is the dot product of the i-th and j-th columns of A. Since A has property 4, we have $< w_i, w_j >= 0$ if $i \neq j$ and $< w_i, w_i >= 1$ for $i = 1, \ldots, n$. This means that $w_1, \ldots, w_n$ is an orthonormal basis. Clearly A is the transition matrix form the basis $w_1, \ldots, w_n$ to the basis $v_1, \ldots, v_n$. This proves that 4 implies 3.

Finally, it is obvious that 3 implies 2. In summary, we proved the implications $2 \Rightarrow 1 \Rightarrow 5 \Rightarrow 4 \Rightarrow 3 \Rightarrow 2$. This means that all the properties 1-5 are equivalent to each other.

Since $({}_E I_D)({}_D I_B) = {}_E I_B$, we easily conclude that product of orthogonal matrices is orthogonal.

Product of orthogonal matrices is orthogonal. The inverse of an orthogonal matrix is orthogonal.

We can also prove this directly: if $AA^t = I = BB^t$ then

$$(AB)(AB)^t = (AB)(B^tA^t) = A(BB^t)A^t = AIA^t = AA^t = I.$$

Example. For any α the matrix $\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$ is an orthogonal matrix. Considered as a linear transformation $\mathbb{R}^2 \longrightarrow \mathbb{R}^2$, this matrix represents rotation counterclockwise about the origin by the angle α.

Exercise. Prove that if A is orthogonal then $\det A = \pm 1$.

Exercise. Prove that if A is a 2×2 orthogonal matrix then either $A = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$ or $A = \begin{bmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{bmatrix}$ for some α.

Orthogonal matrices show up also in another context. First a definition.

Definition. Let V be an inner product space. A linear transformation $T : V \longrightarrow V$ is called an **isometry** if $||Tv|| = ||v||$ for every $v \in V$.

We have the following result.

Theorem. Let $T : V \longrightarrow V$ be a linear transformation from an inner product space to itself. Following properties are equivalent.

1. T is an isometry.
2. $< Tv, Tw >=< v, w >$ for any $v, w \in V$.
3. There is an orthonormal basis $v_1, \ldots, v_n$ of V such that $T(v_1), \ldots, T(v_n)$ is also an orthonormal basis.
4. $T(v_1), \ldots, T(v_n)$ is an orthonormal basis of V for any orthonormal basis $v_1, \ldots, v_n$ of V.
5. There is an orthonormal basis $v_1, \ldots, v_n$ of V in which the matrix representation of T is an orthogonal matrix.
6. The matrix representation of T in any orthonormal basis of V is an orthogonal matrix.

Indeed, suppose that T is an isometry. Recall that $2 < v, w >= ||v + w||^2 - ||v||^2 - ||w||^2$ for any $v, w \in V$. Thus

$$2 < Tv, Tw >= ||Tv+Tw||^2-||Tv||^2-||Tw||^2 = ||T(v+w)||^2-||Tv||^2-||Tw||^2 = ||v+w||^2-||v||^2-||w||^2 = 2 < v, w > .$$

This shows that property 2 holds. Conversely, if property 2 holds then taking $v = w$ we see that $< T(v), T(v) >=< v, v >$ for any v. This clearly is the same as property 1. Thus properties 1 and 2 are equivalent.

Suppose now that property 2 holds. If $v_1, \ldots, v_n$ is an orthonormal basis of V then

$$< Tv_i, Tv_j >=< v_i, v_j >= \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}.$$

Thus $T(v_1), \ldots, T(v_n)$ is an orthonormal basis of V. Thus property 4 holds.

Suppose now that property 4 holds. Pick an orthonormal basis $v_1, \ldots, v_n$. Note that the matrix representation of T in the basis $v_1, \ldots, v_n$ is the same as the transition matrix from the basis $T(v_1), \ldots, T(v_n)$ to the basis $v_1, \ldots, v_n$ (the i-th column of each of these matrices provides coordinates of $T(v_i)$ in the basis $v_1, \ldots, v_n$.). Since both bases $v_1, \ldots, v_n$ and $T(v_1), \ldots, T(v_n)$ are orthonormal, we know that the transition matrix is an orthogonal matrix. This shows that property 6 holds.

Clearly property 5 is a special case of property 6.

Suppose now that property 5 holds. Since the matrix of T in the basis $v_1, \ldots, v_n$ is orthogonal, it is invertible so T is invertible. It follows that $T(v_1), \ldots, T(v_n)$ is a basis of V. As we observed above, the matrix representation of T in the basis $v_1, \ldots, v_n$ is the same as the transition matrix from the basis $T(v_1), \ldots, T(v_n)$ to the basis $v_1, \ldots, v_n$. Thus the transition matrix from the basis $T(v_1), \ldots, T(v_n)$ to the basis $v_1, \ldots, v_n$ is orthogonal and so its its inverse, i.e. the the transition matrix from the basis $v_1, \ldots, v_n$ to the basis $T(v_1), \ldots, T(v_n)$ is orthogonal. We know that an orthogonal matrix transitions an orthogonal basis to an orthogonal basis. Thus $T(v_1), \ldots, T(v_n)$ is an orthogonal basis. This shows that property 3 holds.

Finally, assume property 3. Consider two vectors v, w in V. Clearly the coordinates of v (respectively w) in the basis $v_1, \ldots, v_n$ are the same as the coordinates of Tv (respectively Tw) in the basis $T(v_1), \ldots, T(v_n)$. Recall now that inner product of two vectors is equal to the dot product of the vectors of coordinates in any orthonormal basis. This shows that $< v, w >=< Tv, Tw >$, i.e. property 2 holds.

In summary, we proved the implications $2 \Rightarrow 4 \Rightarrow 6 \Rightarrow 5 \Rightarrow 3 \Rightarrow 2$. We also showed that 1 and 2 are equivalent. This means that all the properties 1-6 are equivalent to each other.

Challenging exercise. Suppose that $f : V \longrightarrow V$ is a function such that $f(0) = 0$ and $||f(v)-f(w)|| = ||v - w||$ for any v, w. Prove that f is an isometry (the key point here is to prove that f must be a linear transformation).

Recall now that a particularly nice linear transformations are those which are diagonalizable. This means that there exists a basis consisting of eigenvectors. In the context of inner product spaces, the analogous property would ask for an orthonormal basis consisting of eigenvectors. Consider then a linear transformation $T : V \longrightarrow V$, where V is an inner product space. Suppose that $v_1, \ldots, v_n$ is an orthonormal basis consisting of eigenvectors. Thus the matrix P of T in this basis is diagonal. Consider another orthonormal basis $w_1, \ldots, w_n$ and let A be the transition matrix from the basis $w_1, \ldots, w_n$ to the basis $v_1, \ldots, v_n$. Then A is orthogonal and $A^{-1}PA$ is the matrix of T in the basis $w_1, \ldots, w_n$. Since $A^{-1} = A^t$ and $P = P^t$ is symmetric, we see that A^tPA is symmetric. Thus the matrix of T in any orthonormal basis is symmetric. We have the following result.

Proposition. Let V be an inner product space and let $T : V \longrightarrow V$ be a linear transformation. Following conditions are equivalent.

1. $< Tv, w > = < v, Tw >$ for any $v, w \in V$.
2. There is an orthonormal basis in which T is represented by a symmetric matrix.
3. The matrix representing T in any orthonormal basis is symmetric.

To justify the proposition, suppose first that property 1 holds. If $v_1, \ldots, v_n$ is an orthonormal basis of V, then the i, j-entry of the matrix M representing T in this basis is $< v_i, Tv_j >$. Since $< v_j, Tv_i > = < Tv_i, v_j > = < v_i, Tv_j >$ (the last equality follows from preoperty 1), the matrix M is symmetric. This shows that property 3 holds.

Clearly property 2 is a special case of property 3.

Finally, suppose that property 2 holds for some orthonormal basis $v_1, \ldots, v_n$. Let M be the matrix representing T in this basis, so M is symmetric. Consider two vectors $v, w \in V$ and let $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ be the coordinates of v and w respectively in the basis $v_1, \ldots, v_n$. Let $A = [a_1, \ldots, a_n]$, $B = [b_1, \ldots, b_n]$ be $1 \times n$ matrices whose row list coordinates of v, w respectively. Then the entries of the $n \times 1$ matrices MA^t, MB^t are the coordinates of Tv, Tw respectively (in the basis $v_1, \ldots, v_n$). The dot product of the vectors of coordinates of v and TW is the only entry of the matrix $A(MB^t)$, so $[< v, Tw >] = A(MB^t)$. Similarly, $[< Tv, w >] = [< w, Tv >] = B(MA^t)$. Clearly, every 1×1 matrix is symmetric. Thus

$$[< v, Tw >] = [< v, Tw >]^t = (A(MB^t)^t = (B^t)^t M^t A^t = BMA^t = [< Tv, w >].$$

This means that $< Tv, w > = < v, Tw >$. So property 1 holds.

In summary, we proved the implications $1 \Rightarrow 3 \Rightarrow 2 \Rightarrow 1$. This means that all the properties 1-3 are equivalent to each other.

Definition. Let V be an inner product space. A linear transformation $T : V \longrightarrow V$ is called an **self-adjoint** if $< Tv, w > = < v, Tw$ for every $v, w \in V$.

The name may seem unmotivated at this point, but there is a good reason for it. Unfortunately, providing more details now would take us too far away. Note that the proposition implies in particular that a matrix A is symmetric if and only if the linear transformation L_A is self-adjoint (when we use the dot product as an inner product on $\mathbb{R}^n$).

Note that we can now reformulate an observation we made earlier as follows:

Let V be an inner product space. If V has an orthonormal basis consisting of eigenvectors of a linear transformation $T : V \longrightarrow V$ then T is self-adjoint. In particular, if $\mathbb{R}^n$ has an orthonormal basis consisting of eigenvectors of a matrix A then A is symmetric.

It turns out that the converse to this statement is also true. In other words, we have the following important theorem.

Symmetric Matrix Theorem. Let V be an inner product space. V has an orthonormal basis consisting of eigenvectors of a linear transformation $T : V \longrightarrow V$ if and only if T is self-adjoint. In particular, an $n \times n$ matrix A is symmetric if and only if $\mathbb{R}^n$ has an orthonormal basis consisting of eigenvectors of A.

We will devote the rest of this note to a proof of this result. The key to proving this result is the following:

A self-adjoint linear transformation has an eigenvalue.

Indeed, suppose that we already proved this result. Consider a self-adjoint linear transformation $T : V \longrightarrow V$, where $\dim V = n$. We know that T has an eigenvalue λ_1. Let v_1 be an eigenvector corresponding to λ_1 and let $W = \{v_1\}^\perp$. Replacing v_1 by $v_1/||v_1||$ if necessary, we may assume $||v_1|| = 1$. Thus W is a subspace of V of dimension $n-1$. Note that for any $w \in W$ we have

$$< v_1, Tw >=< Tv_1, w >=< \lambda_1 v_1, w >= \lambda_1 < v_1, w >= 0.$$

This shows that $Tw \in W$. In other words, T maps W into W. We consider W as an inner product space with the inner product being the restriction of the inner product on V to the subspace W. Then $T_W : W \longrightarrow W$ is clearly self-adjoint, where we write T_W for the restriction of T to W. Using induction, we can assume that we have already constructed an orthonormal basis $v_2, \ldots, v_n$ of W consisting of eigenvectors of T_W. Thus $v_1, \ldots, v_n$ is an orthonormal basis of V consisting of eigenvectors of T. This proves the hard part of the symmetric matrix theorem.

It remains to prove our key statement, that a self-adjoint linear transformation has an eigenvalue. Let V be an inner product space and let $T : V \longrightarrow V$ be a self-adjoint linear transformation. We will consider the set $U = \{v \in V : ||v|| = 1$ of all vectors of length 1. Choosing and orthonormal basis, we see that a vector v is in U if and only if its coordinates $x_1, \ldots, x_n$ satisfy $x_1^2 + \ldots + x_n^2 = 1$. This identifies U with the subset $S_n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1^2 + \ldots + x_n^2 = 1$ of $\mathbb{R}^n$ which is often called the unit n-sphere. We will use without proof the following well known fact from analysis:

Every continuous function $f : S_n \longrightarrow \mathbb{R}$ attains its maximum value at some point of S_n.

This is not a difficult result. It extends the result form calculus (often called extreme value theorem), that a continuous function on a closed interval attains its largest and smallest value. To prove this fact is a good exercise in calculus (it only requires the fact that a bounded sequence of real numbers has a convergent subsequence).

We will apply this to the function $f : U \longrightarrow \mathbb{R}$ given by $f(v) =< v, Tv >$. Expressed in coordinates $x_1, \ldots, x_n$ this is easily seen to be a polynomial function from S_n to $\mathbb{R}$, hence continuous. Thus there us $u \in U$ such that $f(u)$ is the largest value of f on U. Let $W = \{u\}^\perp$. Choose any $w \in W$ such that $||w|| = 1$. Note that for any t the vector $\cos tu + \sin tw$ is in U. Indeed

$$< \cos tu+\sin tw \cos tu+\sin tw, \cos tu+\sin tw >= \cos^2 t < u, u > + \sin^2 t < w, w > +2 \sin t \cos t < u, w > \cos^2 t+\sin^2 t = 1$$

since $< u, u >= 1 =< w, w, >$ and $< u, w >= 0$. Let $g(t) = f(\cos tu + \sin tw)$. Thus $g(0) = f(u)$ and

$$g(t) = f(\cos tu+\sin tw) =< \cos tu+\sin tw, T(\cos tu+\sin tw) >=< \cos tu+\sin tw, \cos tT(u)+\sin tT(w) >$$

$$=< u, Tu > \cos^2 t+ < w, Tw > \sin^2 t + (< u, Tw > + < w, Tu >) \sin t \cos t.$$

Now g is a differentiable function of t and $g(0) = f(u)$ is the largest value of g. This means that $g'(0) = 0$. Since

$$g'(t) = -2 < u, Tu > \cos t \sin t + 2 < w, Tw > \sin t \cos t + (< u, Tw > + < w, Tu >)(\cos^2 t - \sin^2 t)$$

we see that $0 = g'(0) =< u, Tw > + < w, Tu >$. Since T is self adjoin, we have $< u, Tw >=< Tu, w >=< w, Tu >$. It follows that $2 < w, Tu >= 0$. In other words, Tu is orthogonal to every vector w in W of length 1. This implies that that Tu is orthogonal to every vector in W. Indeed, if $v \in W$

then $w = v/||v||$ is in W and has length 1, so $< v, Tu >=< ||v||w, Tu >= ||v|| < w, Tu >= 0$. We see that $Tu \in W^{\perp} = (\{u\}^{\perp})^{\perp} = \text{span}\{u\}$. This means that $Tu = \lambda u$ for some λ. Thus we proved that

u is an eigenvector of T and therefore T has an eigenvalue.

This completes the proof of the symmetric matrix theorem. The fact that the the vector where f attains largest value is an eigenvector of T has important geometric meaning, but we will not discuss it here.

Te symmetric matrix theorem tells us that if A is a symmetric $n \times n$ matrix then there is an orthonormal basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$ consisting of eigenvectors for A. If K is the transition matrix form the basis $v_1, \ldots, v_n$ to the standard basis then K is an orthogonal matrix and $K^{-1}AK$ is diagonal. We often say that A is orthogonally diagonalizable. Of course, we would like to have a procedure allowing us to find K and the orthonormal basis $v_1, \ldots, v_n$. It turns out that we can follow the same procedure we use to diagonalize a matrix, with just a small modification. Here are the steps

- find all eigenvalues of A.
- for every eigenvalue λ find an orthonormal basis of the eigenspace $V(\lambda)$ (this is the small modification, as we want the basis to be orthonormal). We can first find a basis of $V(\lambda)$, then apply Gramm-Schmidt process to get an orthogonal basis, and then normalize the vectors to get an orthonormal basis.
- combine the orthonormal bases for each eigenspace. This will give you an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors for A.

We need to justify the claim in last step. It easily folows from the following result, which is of independent interest.

Let $T : V \longrightarrow V$ be self-adjoint. If v,w are eigenvectors of T corresponding to different eigenvalues, then $v \perp w$.

Indeed, we have $T(v) = av$ and $Tw = bw$ for some scalars $a \neq b$. Now

$$a < v, w >=< av, w >=< Tv, w >=< v, Tw >=< v, bw >= b < v, w >$$

so $(a - b) < v, w >= 0$. Since $a - b \neq 0$, we have $< v, w >= 0$, i.e. $v \perp w$.