Lecture 18 recap

1) Definition of a vector orthogonal to a space. How to find 'all' vectors orthogonal to a space.
2) Definition (and uniqueness) of orthogonal projection.
3) Orthogonal projection theorem. The requirement for orthogonal (orthonormal) bases.
4) Using orthogonal projections to find orthogonal bases - Gram-Schmidt Process.
5) Orthogonal projections and best approximations.
6) What is a least squares solution to a linear system.

LECTURE 19

Best approximation (cont'd)

Orthogonal matrices

Orthogonal diagonalization (Section 6.3)

Theorem (Least squares solution)

Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a linear system. Then $\boldsymbol{x}$ is a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$ if and only if $\boldsymbol{x}$ is a solution to

$$\boxed{\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{A}^T\boldsymbol{b}}$$

Proof: Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{pmatrix}$ $\quad \boldsymbol{u}_i = i$th column of $\boldsymbol{A}$

Let $V = \text{span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\} =$ column space of $\boldsymbol{A}$.

$\boldsymbol{x}$ is a least squares solution

$\Leftrightarrow$ $\boldsymbol{x}$ is a solution to $\boldsymbol{Ax} = \boldsymbol{p}$ $\quad$ ($\boldsymbol{p}$ is the projection of $\boldsymbol{b}$ onto V)

$\Leftrightarrow$ $\boldsymbol{Ax}$ is the projection of $\boldsymbol{b}$ onto V

Theorem (Least squares solution)

Proof: Let $\boldsymbol{A} = (\boldsymbol{u}_1 \quad \boldsymbol{u}_2 \quad \dots \quad \boldsymbol{u}_n)$ $\quad \boldsymbol{u}_i = i$th column of $\boldsymbol{A}$

Let $V = \text{span}\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\} =$ column space of $\boldsymbol{A}$.

$\boldsymbol{x}$ is a least squares solution

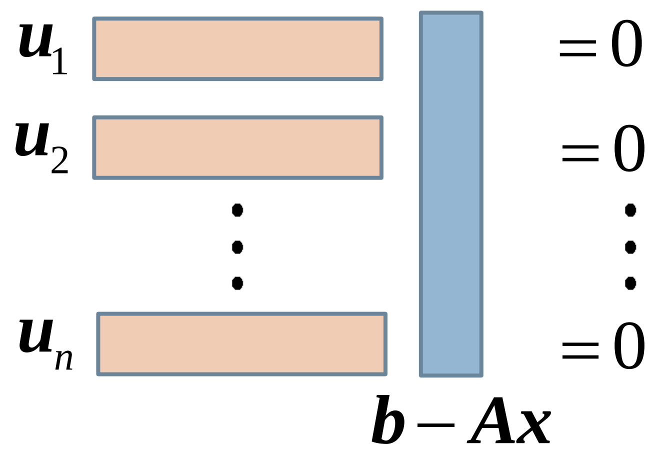
$\Leftrightarrow$ $\boldsymbol{x}$ is a solution to $\boldsymbol{Ax} = \boldsymbol{p}$ $\quad$ ($\boldsymbol{p}$ is the projection of $\boldsymbol{b}$ onto V)

$\Leftrightarrow$ $\boldsymbol{Ax}$ is the projection of $\boldsymbol{b}$ onto V

$\Leftrightarrow$ $\boldsymbol{b} - \boldsymbol{Ax}$ is orthogonal to V

$\Leftrightarrow$ $\boldsymbol{u}_i \cdot (\boldsymbol{b} - \boldsymbol{Ax}) = 0$ for all $i = 1, \dots, n$.

$\Leftrightarrow$ $A^T(\boldsymbol{b} - \boldsymbol{Ax}) = \boldsymbol{0}$ $\Leftrightarrow$ $\boxed{\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{A}^T\boldsymbol{b}}$



Back to our experiments!

i	1	2	3	4	5	6
r_i	0	0	1	1	2	2
s_i	0	1	2	0	1	2
t_i	0.5	1.6	2.8	0.8	5.1	5.9

$$\begin{pmatrix} r_1^2 & s_1 & 1 \\ r_2^2 & s_2 & 1 \\ \vdots & \vdots & \vdots \\ r_6^2 & s_6 & 1 \end{pmatrix} \begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_6 \end{pmatrix} \Leftrightarrow \boldsymbol{Ax} = \boldsymbol{b} \qquad \boldsymbol{A} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 4 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix} \qquad \boldsymbol{b} = \begin{pmatrix} 0.5 \\ 1.6 \\ 2.8 \\ 0.8 \\ 5.1 \\ 5.9 \end{pmatrix}$$

Back to our experiments!

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{pmatrix} 34 & 14 & 10 \\ 14 & 10 & 6 \\ 10 & 6 & 6 \end{pmatrix} \qquad \boldsymbol{A}^T\boldsymbol{b} = \begin{pmatrix} 47.6 \\ 24.1 \\ 16.7 \end{pmatrix}$$

Solving

$$\begin{pmatrix} 34 & 14 & 10 \\ 14 & 10 & 6 \\ 10 & 6 & 6 \end{pmatrix} \begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} 47.6 \\ 24.1 \\ 16.7 \end{pmatrix}$$

we have $c = 0.9275$, $d = 0.9225$, $e = 0.3150$.

Example

Let $V = \text{span}\{(1,-1,1,-1),(1,2,0,1),(2,1,1,0)\}$. Find the projection of $(1,1,1,1)$ onto V.

Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.

Note that V is the column space of $\boldsymbol{A}$.

We first obtain a least squares solution of $\boldsymbol{Ax} = \boldsymbol{b}$.

Example

Let $V = \text{span}\{(1,-1,1,-1),(1,2,0,1),(2,1,1,0)\}$. Find the projection of $(1,1,1,1)$ onto V.

We first obtain a least squares solution of $\boldsymbol{Ax} = \boldsymbol{b}$.

Solving $\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{A}^T\boldsymbol{b}$:

$$\begin{pmatrix} 4 & -2 & 2 \\ -2 & 6 & 4 \\ 2 & 4 & 6 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 4 \end{pmatrix}$$

solving...

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -t + \frac{2}{5} \\ -t + \frac{4}{5} \\ t \end{pmatrix}, t \in \mathbb{R}$$

Infinitely many least squares solutions

choose one: $\begin{pmatrix} \frac{2}{5} \\ \frac{4}{5} \\ 0 \end{pmatrix}$

Example

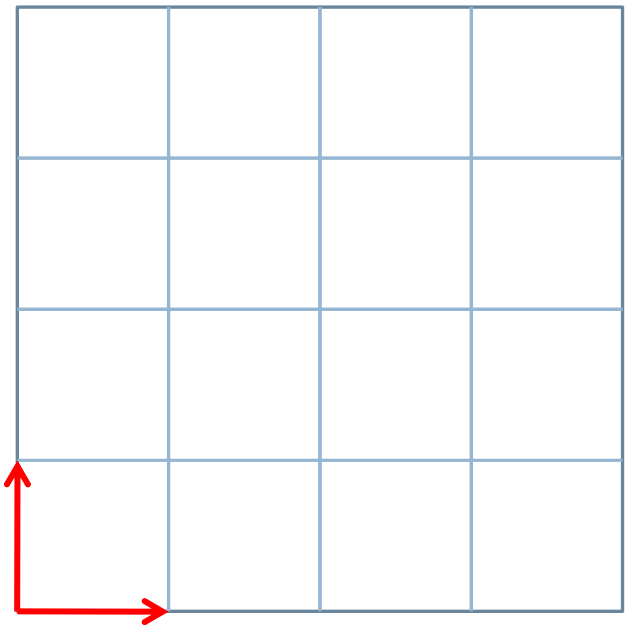
Let $V = \text{span}\{(1,-1,1,-1),(1,2,0,1),(2,1,1,0)\}$. Find the projection of $(1,1,1,1)$ onto V.

$$\boldsymbol{A}\boxed{\begin{pmatrix} \frac{2}{5} \\ \frac{4}{5} \\ 0 \end{pmatrix}} = \begin{pmatrix} \frac{6}{5} \\ \frac{6}{5} \\ \frac{2}{5} \\ \frac{2}{5} \end{pmatrix}$$ is the projection of $(1,1,1,1)$ onto V.

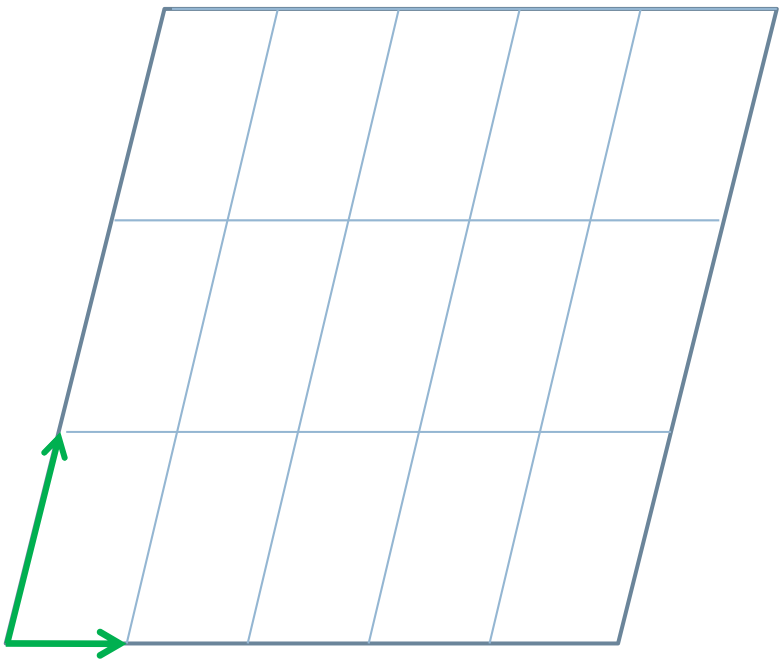
choose one: $\boxed{\begin{pmatrix} \frac{2}{5} \\ \frac{4}{5} \\ 0 \end{pmatrix}}$

Orthonormal bases

Coordinate systems built on orthonormal bases are convenient and has advantages.



orthonormal basis

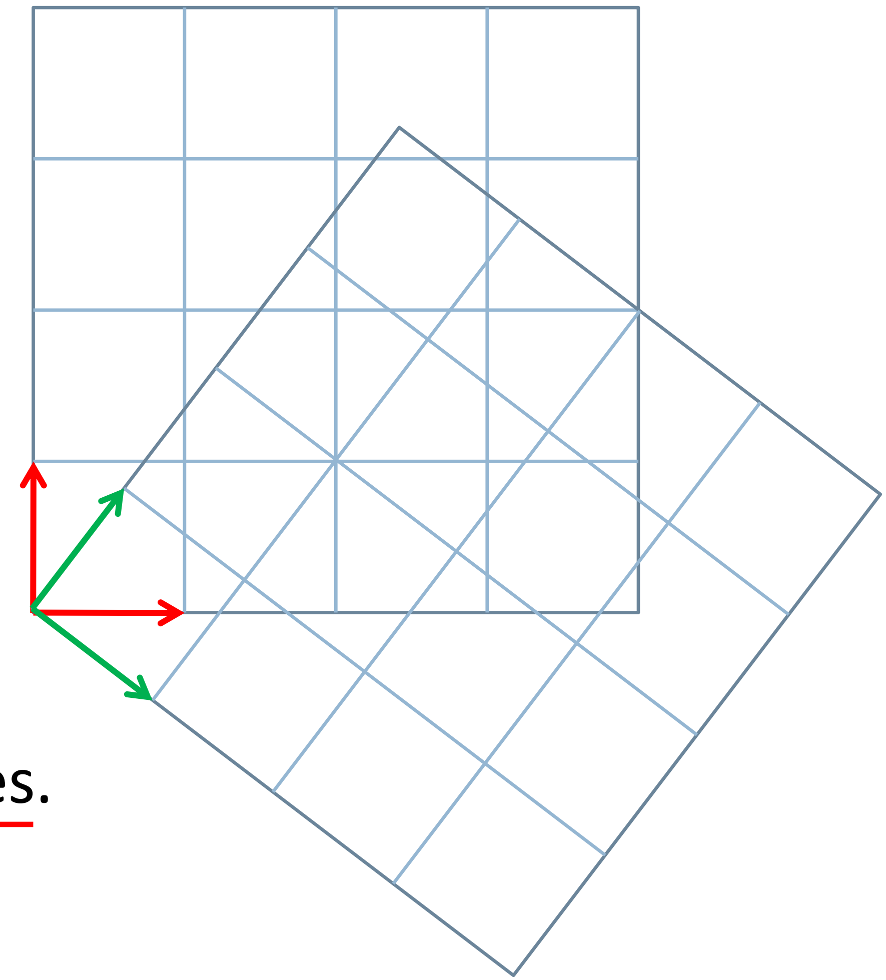


not orthonormal basis

Orthonormal bases

At the same time, it is often necessary to shift between bases when we study vector spaces.

We will now study the transition matrix arising between two orthonormal bases.

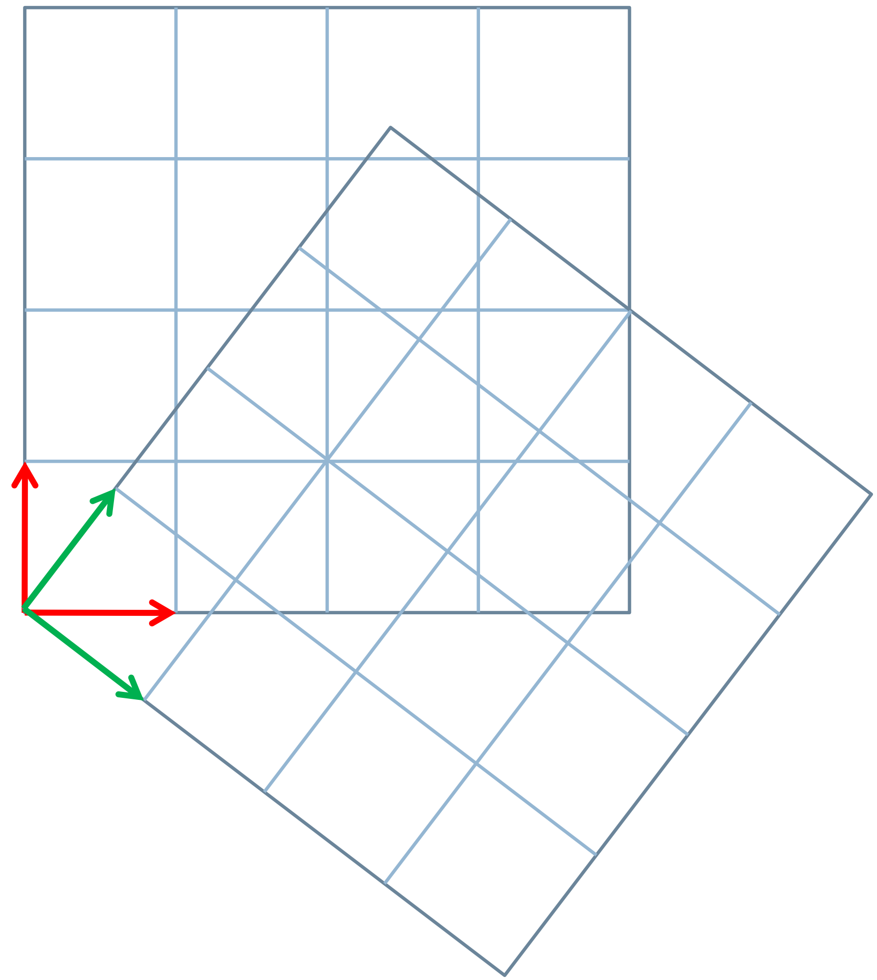


Orthonormal bases

Recall that if $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ and $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ are two orthonormal bases for a vector space V, then the transition matrix from S to T is

$$\boldsymbol{P} = \left([\boldsymbol{u}_1]_T \quad [\boldsymbol{u}_2]_T \quad \cdots \quad [\boldsymbol{u}_k]_T \right)$$

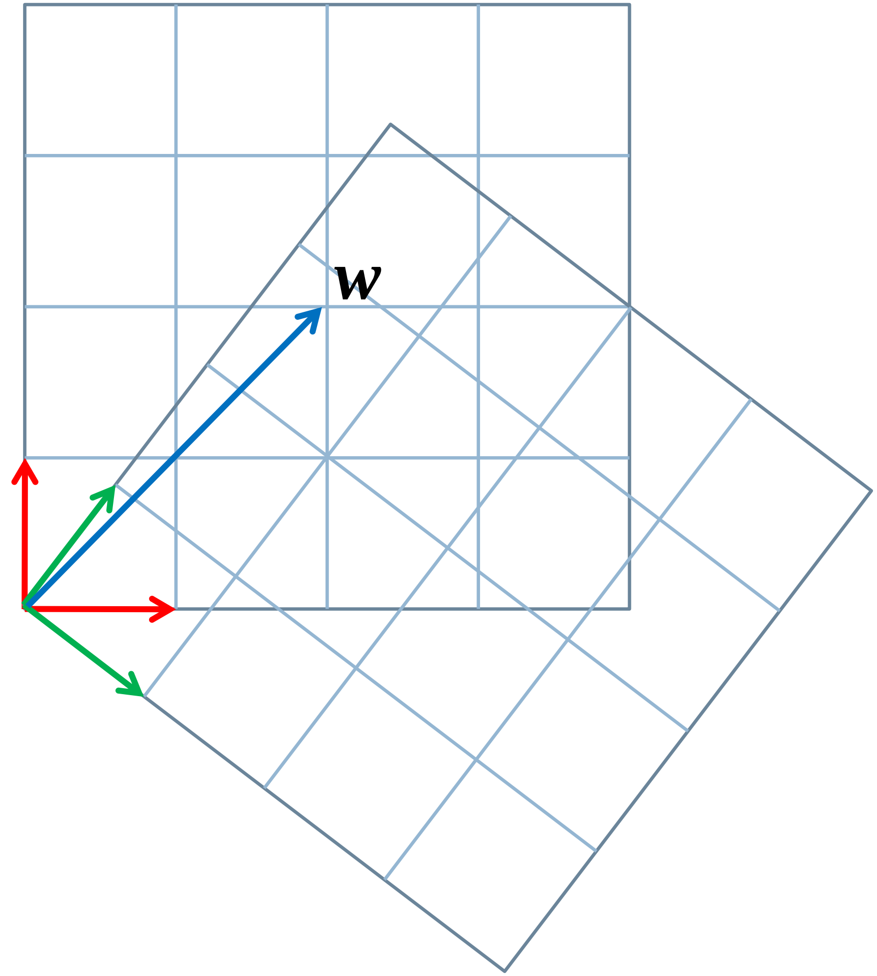
(column) coordinate vector of $\boldsymbol{u}_1$ (from S) relative to T



Orthonormal bases

Then for all vectors $\boldsymbol{w} \in V$,

$$\boldsymbol{P}[\boldsymbol{w}]_S = [\boldsymbol{w}]_T$$



$$\boldsymbol{P} = \left([\boldsymbol{u}_1]_T \quad [\boldsymbol{u}_2]_T \quad \cdots \quad [\boldsymbol{u}_k]_T \right)$$

(column) coordinate vector

of $\boldsymbol{u}_1$ (from S) relative to T

Example (transition matrix between two orthonormal bases)

Let $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ be the standard basis for $\mathbb{R}^3$ and $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ where

$$\boldsymbol{e}_1 = (1,0,0)$$
$$\boldsymbol{e}_2 = (0,1,0)$$
$$\boldsymbol{e}_3 = (0,0,1)$$

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}(1,1,1) \quad \boldsymbol{u}_2 = \frac{1}{\sqrt{2}}(1,0,-1) \quad \boldsymbol{u}_3 = \frac{1}{\sqrt{6}}(1,-2,1)$$

E is an orthonormal basis for $\mathbb{R}^3$.

$\boldsymbol{u}_i \cdot \boldsymbol{u}_j = 0$ for all $i \neq j$ $\quad \|\boldsymbol{u}_i\| = 1$ for all $i = 1,2,3.$

S is also an orthonormal basis for $\mathbb{R}^3$.

What is the transition matrix from S to E?

Need to write vectors in S in terms of vectors in E.

Example (transition matrix between two orthonormal bases)

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}(1,1,1) = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = \boxed{\frac{1}{\sqrt{3}}}\boldsymbol{e}_1 + \boxed{\frac{1}{\sqrt{3}}}\boldsymbol{e}_2 + \boxed{\frac{1}{\sqrt{3}}}\boldsymbol{e}_3$$

$$\boldsymbol{u}_2 = \frac{1}{\sqrt{2}}(1,0,-1) = \left(\frac{1}{\sqrt{2}}, 0, \frac{-1}{\sqrt{2}}\right) = \boxed{\frac{1}{\sqrt{2}}}\boldsymbol{e}_1 + \boxed{0}\boldsymbol{e}_2 \boxed{-\frac{1}{\sqrt{2}}}\boldsymbol{e}_3$$

$$\boldsymbol{u}_3 = \frac{1}{\sqrt{6}}(1,-2,1) = \left(\frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right) = \boxed{\frac{1}{\sqrt{6}}}\boldsymbol{e}_1 \boxed{-\frac{2}{\sqrt{6}}}\boldsymbol{e}_2 + \boxed{\frac{1}{\sqrt{6}}}\boldsymbol{e}_3$$

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$ is the transition matrix from S to E.

Example (transition matrix between two orthonormal bases)

What about the transition matrix from E to S?

$$\boldsymbol{e}_1 = (\boxed{\boldsymbol{e}_1 \cdot \boldsymbol{u}_1})\boldsymbol{u}_1 + (\boxed{\boldsymbol{e}_1 \cdot \boldsymbol{u}_2})\boldsymbol{u}_2 + (\boxed{\boldsymbol{e}_1 \cdot \boldsymbol{u}_3})\boldsymbol{u}_3$$

Since $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ is an orthonormal basis, computing the coefficients is easy.

Need to write vectors in E in terms of vectors in S.

$$\boldsymbol{e}_2 = (\boxed{\boldsymbol{e}_2 \cdot \boldsymbol{u}_1})\boldsymbol{u}_1 + (\boxed{\boldsymbol{e}_2 \cdot \boldsymbol{u}_2})\boldsymbol{u}_2 + (\boxed{\boldsymbol{e}_2 \cdot \boldsymbol{u}_3})\boldsymbol{u}_3$$

$$\boldsymbol{e}_3 = (\boxed{\boldsymbol{e}_3 \cdot \boldsymbol{u}_1})\boldsymbol{u}_1 + (\boxed{\boldsymbol{e}_3 \cdot \boldsymbol{u}_2})\boldsymbol{u}_2 + (\boxed{\boldsymbol{e}_3 \cdot \boldsymbol{u}_3})\boldsymbol{u}_3$$

Example (transition matrix between two orthonormal bases)

What about the transition matrix from E to S?

$$\boldsymbol{e}_1 = \left(\boxed{\frac{1}{\sqrt{3}}}\right)\boldsymbol{u}_1 + \left(\boxed{\frac{1}{\sqrt{2}}}\right)\boldsymbol{u}_2 + \left(\boxed{\frac{1}{\sqrt{6}}}\right)\boldsymbol{u}_3$$

$$\boldsymbol{e}_2 = \left(\boxed{\frac{1}{\sqrt{3}}}\right)\boldsymbol{u}_1 + \left(\boxed{0}\right)\boldsymbol{u}_2 + \left(\boxed{\frac{-2}{\sqrt{6}}}\right)\boldsymbol{u}_3$$

$$\boldsymbol{e}_3 = \left(\boxed{\frac{1}{\sqrt{3}}}\right)\boldsymbol{u}_1 + \left(\boxed{\frac{-1}{\sqrt{2}}}\right)\boldsymbol{u}_2 + \left(\boxed{\frac{1}{\sqrt{6}}}\right)\boldsymbol{u}_3$$

Need to write vectors in E in terms of vectors in S.

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

is the transition matrix from E to S.

Example (transition matrix between two orthonormal bases)

Transition matrix from S to E

$$\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

Transition matrix from E to S

$$\boldsymbol{Q} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

Observe that

1) $\boldsymbol{Q} = \boldsymbol{P}^T$

But we also know that

2) $\boldsymbol{Q} = \boldsymbol{P}^{-1}$ (from Chapter 3)

So $\boxed{\boldsymbol{P}^T = \boldsymbol{P}^{-1}}$ for this example. Is this a coincidence?

Definition (orthogonal matrix)

A square matrix $\boldsymbol{A}$ is said to be orthogonal if

$$\boxed{\boldsymbol{A}^{-1} = \boldsymbol{A}^{T}}$$

(orthogonal matrices must be invertible)

'orthogonal' can be used to describe

between two vectors $\boldsymbol{u},\boldsymbol{v}$

$\boldsymbol{u} \cdot \boldsymbol{v} = 0$

a set S

between a vector $\boldsymbol{u}$ and a space V

$\boldsymbol{u} \cdot \boldsymbol{v} = 0 \ \forall \boldsymbol{v} \in V$

a matrix $\boldsymbol{A}$

How to check if a matrix is orthogonal?

A square matrix $\boldsymbol{A}$ is said to be orthogonal if

$$\boxed{\boldsymbol{A}^{-1} = \boldsymbol{A}^{T}}$$

(orthogonal matrices must be invertible)

By 'half-price' theorem, it suffices to check

$$\boldsymbol{A}\boldsymbol{A}^{T} = \boldsymbol{I} \qquad \text{or} \qquad \boldsymbol{A}^{T}\boldsymbol{A} = \boldsymbol{I}$$

Example (orthogonal matrices)

The following are example of orthogonal matrices.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

Orthogonal are 'special' invertible matrices

Remember that orthogonal matrices are 'special' invertible matrices where $\boldsymbol{A}^{-1} = \boldsymbol{A}^{T}$.

Recall also that we have the following:

$\boldsymbol{A}$ is an invertible square matrix of order n

$\Leftrightarrow$ the rows of $\boldsymbol{A}$ forms a basis for $\mathbb{R}^n$

$\Leftrightarrow$ the columns of $\boldsymbol{A}$ forms a basis for $\mathbb{R}^n$

What about...

$\boldsymbol{A}$ is an orthogonal square matrix of order n

$\Leftrightarrow$??

If $\boldsymbol{A}$ is 'special' will the rows or columns of $\boldsymbol{A}$ be special?

Theorem (equivalent statements)

Let $\boldsymbol{A}$ be a square matrix of order n. The following statements are equivalent.

1) $\boldsymbol{A}$ is orthogonal
2) The rows of $\boldsymbol{A}$ forms an orthonormal basis for $\mathbb{R}^n$.
3) The columns of $\boldsymbol{A}$ forms an orthonormal basis for $\mathbb{R}^n$.

We will now prove that this is not a coincidence…

Transition matrix from S to E

$$\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

Transition matrix from E to S

$$\boldsymbol{Q} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

So $\boldsymbol{P}^T = \boldsymbol{P}^{-1}$ for this example (that is, $\boldsymbol{P}$ is orthogonal).

If $\boldsymbol{P}$ is the transition matrix between a pair of orthonormal bases for some vector space V, will $\boldsymbol{P}$ always be an orthogonal matrix?

Theorem (transition matrix between orthonormal bases)

Let S and T be two orthonormal bases for a vector space and let $\boldsymbol{P}$ be the transition matrix from S to T. Then $\boldsymbol{P}$ is an orthogonal matrix (that is, $\boldsymbol{P}^{-1} = \boldsymbol{P}^T$ is the transition matrix from T to S).

Proof:

Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ and $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$.

Since T is an orthonormal basis,

$$\boldsymbol{u}_1 = (\boldsymbol{u}_1 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_1 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + \ldots + (\boldsymbol{u}_1 \cdot \boldsymbol{v}_k)\boldsymbol{v}_k$$

$$[\boldsymbol{u}_1]_T = \begin{pmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{v}_1 \\ \boldsymbol{u}_1 \cdot \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{u}_1 \cdot \boldsymbol{v}_k \end{pmatrix}$$

Theorem (transition matrix between orthonormal bases)

Proof: Let $S=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_k\}$ and $T=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_k\}$.

Since T is an orthonormal basis,

$$\boldsymbol{u}_1=(\boldsymbol{u}_1\cdot\boldsymbol{v}_1)\boldsymbol{v}_1+(\boldsymbol{u}_1\cdot\boldsymbol{v}_2)\boldsymbol{v}_2+\ldots+(\boldsymbol{u}_1\cdot\boldsymbol{v}_k)\boldsymbol{v}_k$$

$$[\boldsymbol{u}_1]_T=\begin{pmatrix}\boldsymbol{u}_1\cdot\boldsymbol{v}_1\\ \boldsymbol{u}_1\cdot\boldsymbol{v}_2\\ \vdots\\ \boldsymbol{u}_1\cdot\boldsymbol{v}_k\end{pmatrix}$$

$$\boldsymbol{u}_2=(\boldsymbol{u}_2\cdot\boldsymbol{v}_1)\boldsymbol{v}_1+(\boldsymbol{u}_2\cdot\boldsymbol{v}_2)\boldsymbol{v}_2+\ldots+(\boldsymbol{u}_2\cdot\boldsymbol{v}_k)\boldsymbol{v}_k$$

$$[\boldsymbol{u}_2]_T=\begin{pmatrix}\boldsymbol{u}_2\cdot\boldsymbol{v}_1\\ \boldsymbol{u}_2\cdot\boldsymbol{v}_2\\ \vdots\\ \boldsymbol{u}_2\cdot\boldsymbol{v}_k\end{pmatrix}$$

Transition matrix from S to T

$$\boldsymbol{P}=\begin{pmatrix}[\boldsymbol{u}_1]_T & [\boldsymbol{u}_2]_T & \cdots & [\boldsymbol{u}_k]_T\end{pmatrix}$$

Theorem (transition matrix between orthonormal bases)

Proof: Let $S=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_k\}$ and $T=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_k\}$.

Transition matrix from S to T

$$\boldsymbol{P}=\begin{pmatrix} \boldsymbol{u}_1\cdot\boldsymbol{v}_1 & \boldsymbol{u}_2\cdot\boldsymbol{v}_1 & \cdots & \boldsymbol{u}_k\cdot\boldsymbol{v}_1 \\ \boldsymbol{u}_1\cdot\boldsymbol{v}_2 & \boldsymbol{u}_2\cdot\boldsymbol{v}_2 & \cdots & \boldsymbol{u}_k\cdot\boldsymbol{v}_2 \\ \vdots & \vdots & & \vdots \\ \boldsymbol{u}_1\cdot\boldsymbol{v}_k & \boldsymbol{u}_2\cdot\boldsymbol{v}_k & \cdots & \boldsymbol{u}_k\cdot\boldsymbol{v}_k \end{pmatrix}$$

$$[\boldsymbol{u}_1]_T=\begin{pmatrix} \boxed{\boldsymbol{u}_1\cdot\boldsymbol{v}_1} \\ \boxed{\boldsymbol{u}_1\cdot\boldsymbol{v}_2} \\ \vdots \\ \boxed{\boldsymbol{u}_1\cdot\boldsymbol{v}_k} \end{pmatrix}$$

$$[\boldsymbol{u}_2]_T=\begin{pmatrix} \boxed{\boldsymbol{u}_2\cdot\boldsymbol{v}_1} \\ \boxed{\boldsymbol{u}_2\cdot\boldsymbol{v}_2} \\ \vdots \\ \boxed{\boldsymbol{u}_2\cdot\boldsymbol{v}_k} \end{pmatrix}$$

Theorem (transition matrix between orthonormal bases)

Proof: Let $S=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_k\}$ and $T=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_k\}$.

Transition matrix from S to T

$$\boldsymbol{P}=\begin{pmatrix} \boldsymbol{u}_1\cdot\boldsymbol{v}_1 & \boxed{\boldsymbol{u}_2\cdot\boldsymbol{v}_1} & \cdots & \boldsymbol{u}_k\cdot\boldsymbol{v}_1 \\ \boldsymbol{u}_1\cdot\boldsymbol{v}_2 & \boldsymbol{u}_2\cdot\boldsymbol{v}_2 & \cdots & \boldsymbol{u}_k\cdot\boldsymbol{v}_2 \\ \vdots & \vdots & & \vdots \\ \boxed{\boldsymbol{u}_1\cdot\boldsymbol{v}_k} & \boldsymbol{u}_2\cdot\boldsymbol{v}_k & \cdots & \boldsymbol{u}_k\cdot\boldsymbol{v}_k \end{pmatrix}$$

Note that $\boxed{\boldsymbol{u}_i\cdot\boldsymbol{v}_j}=\boxed{\boldsymbol{v}_j\cdot\boldsymbol{u}_i}$

So $\boldsymbol{Q}=\boldsymbol{P}^T$.

Similarly, we can find the transition matrix from T to S to be

$$\boldsymbol{Q}=\begin{pmatrix} \boldsymbol{v}_1\cdot\boldsymbol{u}_1 & \boldsymbol{v}_2\cdot\boldsymbol{u}_1 & \cdots & \boxed{\boldsymbol{v}_k\cdot\boldsymbol{u}_1} \\ \boxed{\boldsymbol{v}_1\cdot\boldsymbol{u}_2} & \boldsymbol{v}_2\cdot\boldsymbol{u}_2 & \cdots & \boldsymbol{v}_k\cdot\boldsymbol{u}_2 \\ \vdots & \vdots & & \vdots \\ \boldsymbol{v}_1\cdot\boldsymbol{u}_k & \boldsymbol{v}_2\cdot\boldsymbol{u}_k & \cdots & \boldsymbol{v}_k\cdot\boldsymbol{u}_k \end{pmatrix}$$

Theorem (transition matrix between orthonormal bases)

Proof: Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ and $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$.

Transition matrix from S to T $= \boldsymbol{P}$

Transition matrix from T to S $= \boldsymbol{Q}$

So $\boldsymbol{Q} = \boldsymbol{P}^T$.

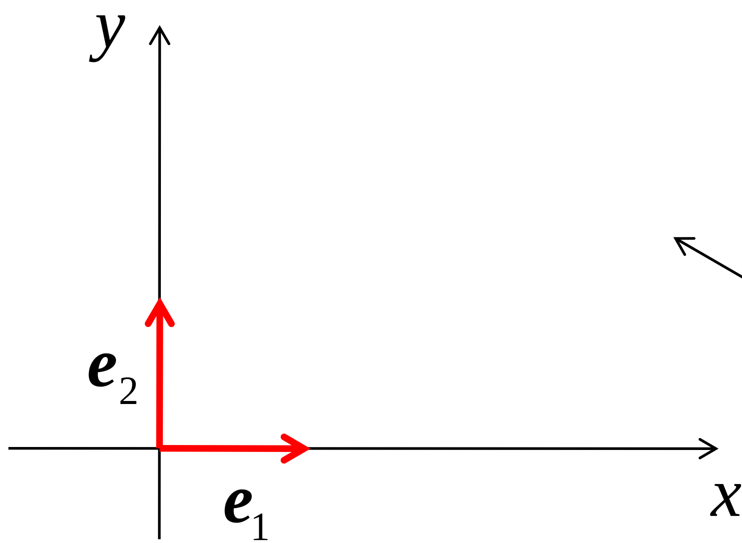
But we also know that $\boldsymbol{Q} = \boldsymbol{P}^{-1}$, so we conclude that

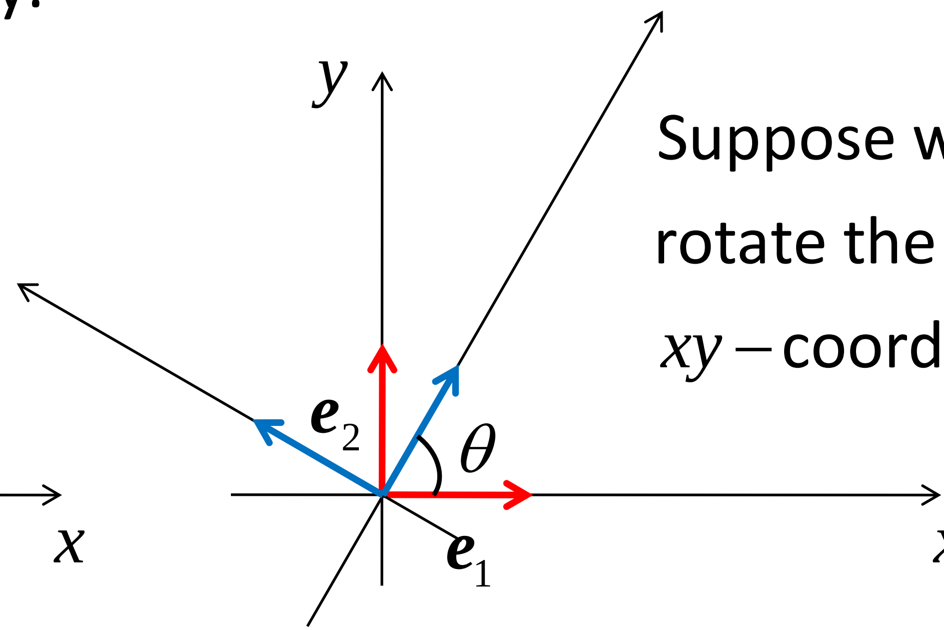
$\boldsymbol{P}^T = \boldsymbol{P}^{-1}$, that is, $\boldsymbol{P}$ is an orthogonal matrix.

Example (rotation of *xy*-coordinates)

Consider the standard basis $\{\underset{\boldsymbol{e}_1}{(1,0)},\underset{\boldsymbol{e}_2}{(0,1)}\}$ of $\mathbb{R}^2$.

Note that $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are in the directions of the $x-$axis and $y-$axis respectively.





Suppose we rotate the $xy-$coordinates

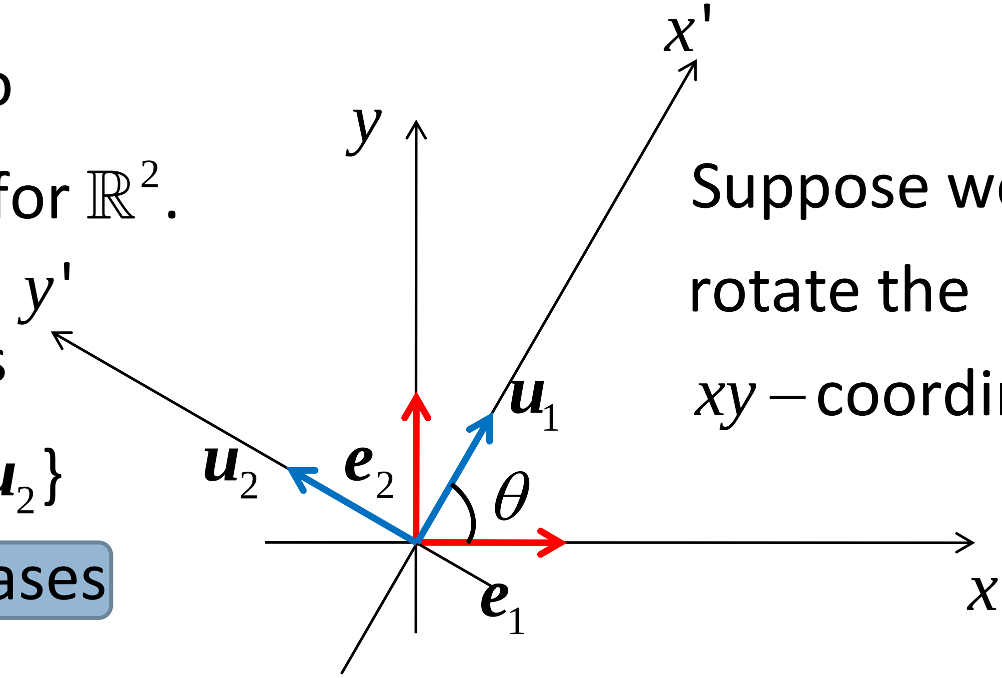
Example (rotation of *xy*-coordinates)

We now have a new $x'y'-$coordinate system.

Let $\boxed{\boldsymbol{u}_1}$ and $\boxed{\boldsymbol{u}_2}$ be unit vectors along the new axes x' and y' respectively.

Then $S=\{\boldsymbol{u}_1,\boldsymbol{u}_2\}$ is also an orthonormal basis for $\mathbb{R}^2$.

orthogonal unit vectors



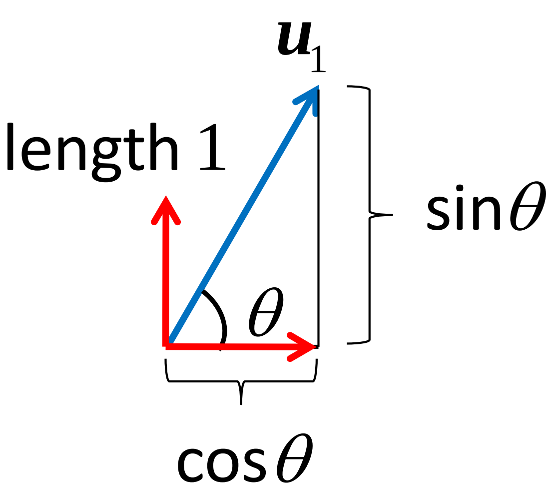
Suppose we rotate the $xy-$coordinates

$E=\{\boldsymbol{e}_1,\boldsymbol{e}_2\}$ and $S=\{\boldsymbol{u}_1,\boldsymbol{u}_2\}$ are two orthonormal bases for $\mathbb{R}^2$.

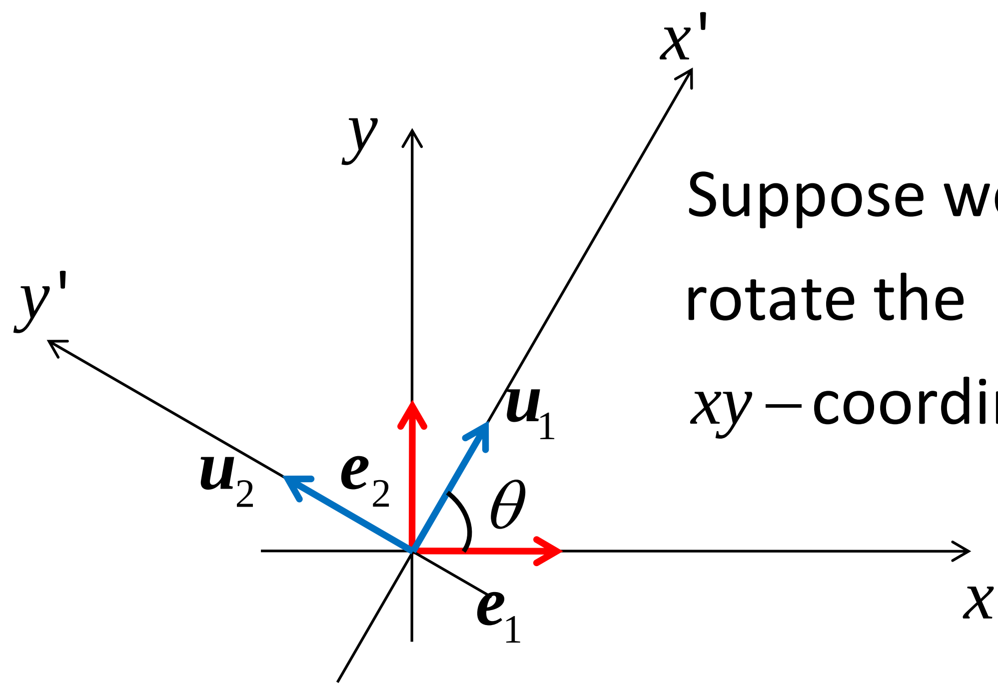
Example (rotation of *xy*-coordinates)

Let us find the transition matrix from S to E.

$E = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ and $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2\}$



$$\boldsymbol{u}_1 = (\cos\theta)\boldsymbol{e}_1 + (\sin\theta)\boldsymbol{e}_2$$

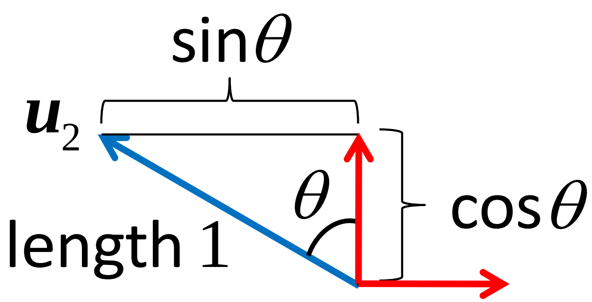


Suppose we rotate the $xy-$coordinates

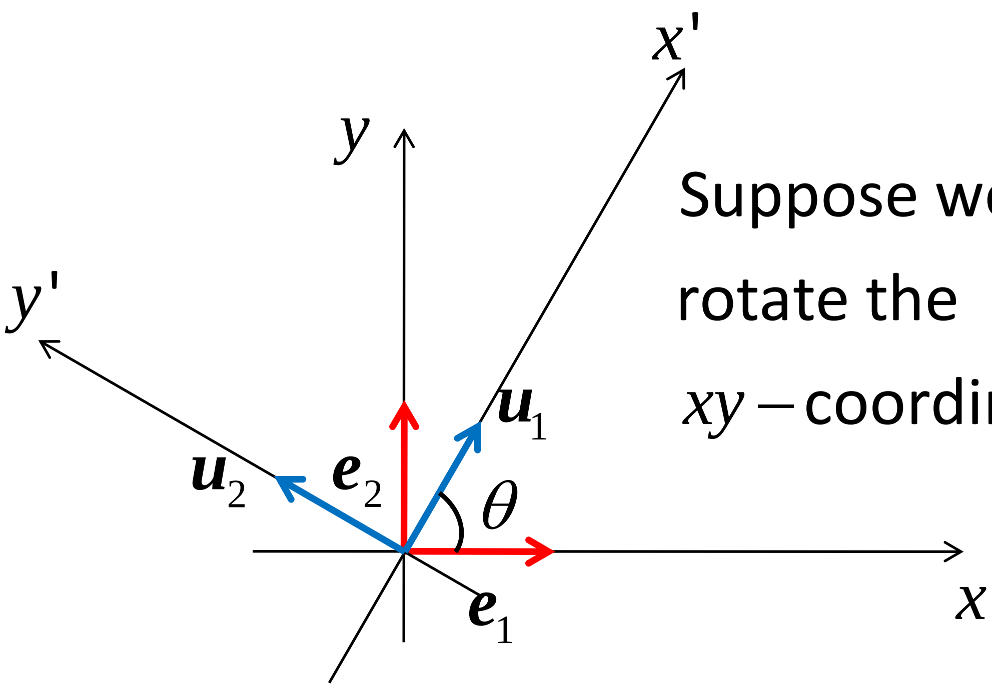
Example (rotation of *xy*-coordinates)

Let us find the transition matrix from S to E.

$E = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ and $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2\}$



$\boldsymbol{u}_2 = (-\sin\theta)\boldsymbol{e}_1 + (\cos\theta)\boldsymbol{e}_2$



Suppose we rotate the $xy-$coordinates

Example (rotation of *xy*-coordinates)

Let us find the transition matrix from S to E.

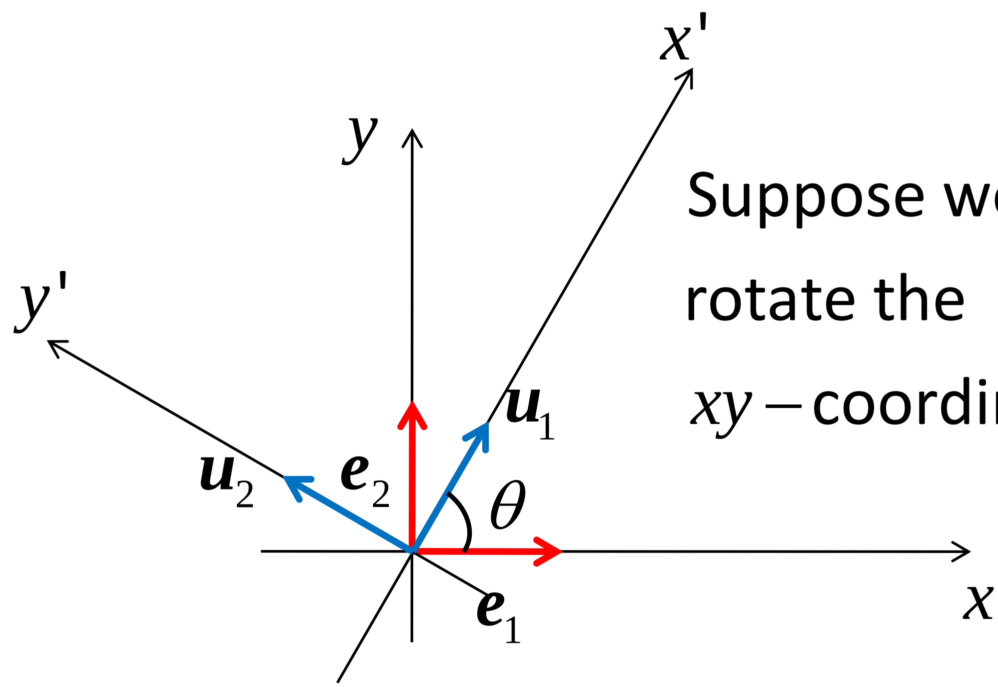
$$\boldsymbol{u}_1 = (\cos\theta)\boldsymbol{e}_1 + (\sin\theta)\boldsymbol{e}_2 \qquad \boldsymbol{u}_2 = (-\sin\theta)\boldsymbol{e}_1 + (\cos\theta)\boldsymbol{e}_2$$

Transition matrix from S to E

$$= \boldsymbol{P} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

$\boldsymbol{P}$ is orthogonal and the transition matrix from E to S is

$$\boldsymbol{P}^T = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$



Suppose we rotate the $xy-$coordinates

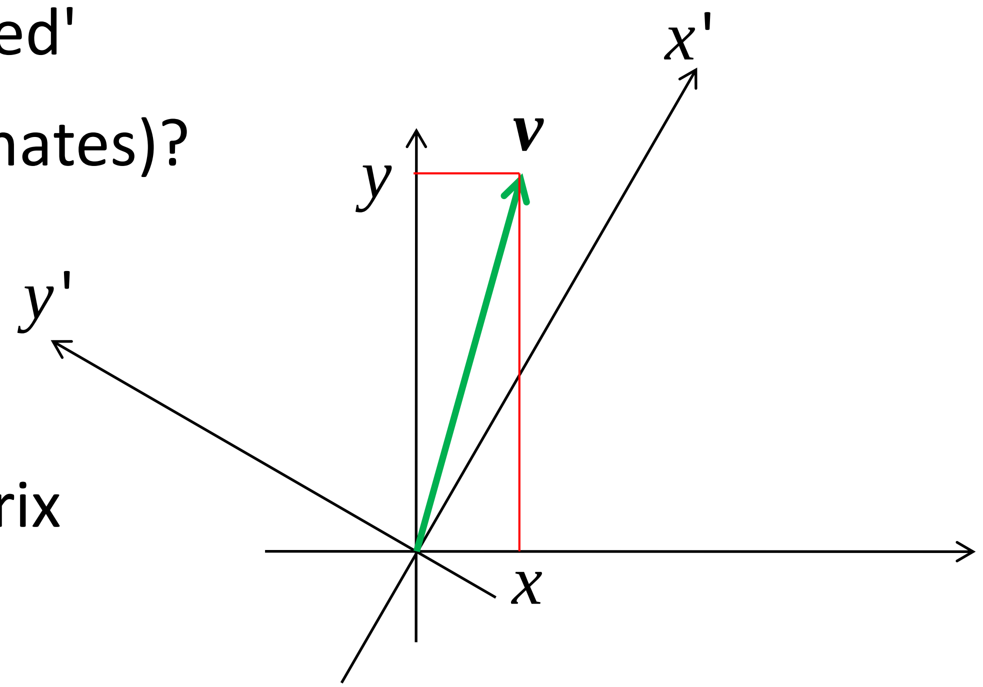
Example (rotation of *xy*-coordinates)

Let $\boldsymbol{v} = (x, y)$ be a vector in $\mathbb{R}^2$, so $[\boldsymbol{v}]_E = \begin{pmatrix} x \\ y \end{pmatrix}$ ('measured' using the $xy-$coordinates).

What is $[\boldsymbol{v}]_S$ ('measured' using the $x'y'-$coordinates)?

$$[\boldsymbol{v}]_S = \boldsymbol{P}^T [\boldsymbol{v}]_E$$

transition matrix from E to S



Example (rotation of *xy*-coordinates)

Let $\boldsymbol{v} = (x, y)$ be a vector in $\mathbb{R}^2$, so $[\boldsymbol{v}]_E = \begin{pmatrix} x \\ y \end{pmatrix}$ ('measured' using the $xy-$coordinates).

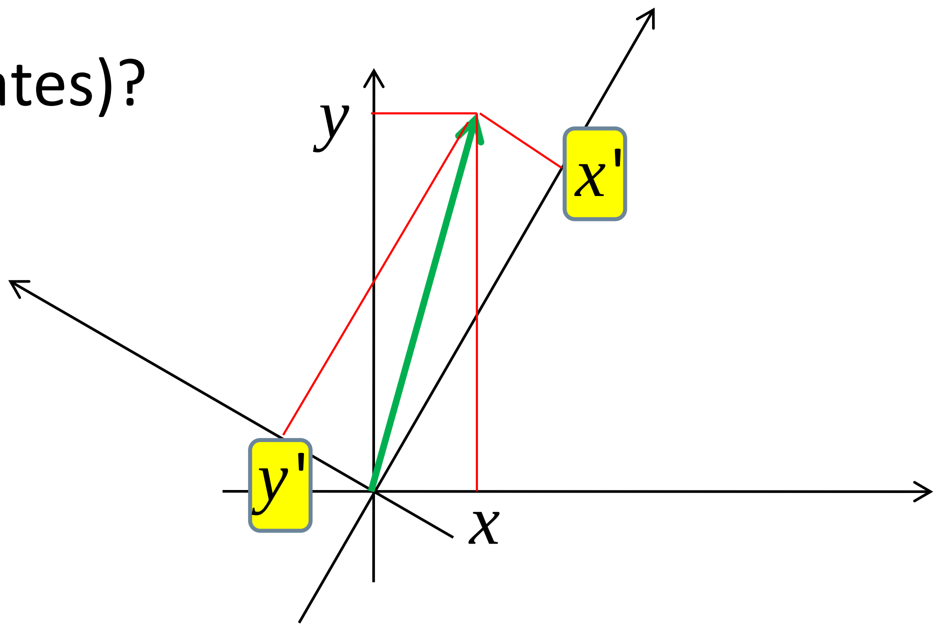
What is $[\boldsymbol{v}]_S$ ('measured' using the $x'y'-$coordinates)?

$$[\boldsymbol{v}]_S = \boldsymbol{P}^T [\boldsymbol{v}]_E$$

$$= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

$$= \begin{pmatrix} x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta \end{pmatrix}$$

$$x' = x\cos\theta + y\sin\theta$$
$$y' = -x\sin\theta + y\cos\theta$$



Recall diagonalization

A square matrix $\boldsymbol{A}$ is said to be diagonalizable if there exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \lambda_1 & & & \mathbf{0} \\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_k \end{pmatrix}$$

Eigenvalues of $\boldsymbol{A}$

To check whether $\boldsymbol{A}$ is diagonalizable we need to determine if we have the required number (equal to the order of $\boldsymbol{A}$) of linearly independent eigenvectors.

Recall diagonalization example

$$\boldsymbol{B} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Recall that 3 and 0 are the eigenvalues of $\boldsymbol{B}$ and

$$E_3 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} \quad E_0 = \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

(G-S Process)

$$= \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \right\}$$

Let $\boldsymbol{R} = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}$.

Note that $\boldsymbol{R}$ is an orthogonal matrix.

That is, $\boldsymbol{R}^T = \boldsymbol{R}^{-1}$. Thus $\boldsymbol{R}^T\boldsymbol{B}\boldsymbol{R} = \boldsymbol{R}^{-1}\boldsymbol{B}\boldsymbol{R} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Definition (orthogonal diagonalization)

A square matrix $\boldsymbol{A}$ is said to be diagonalizable if there exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{AP}$ is a diagonal matrix.

$$\boldsymbol{P}^{T}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \lambda_1 & & & \mathbf{0} \\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_k \end{pmatrix}$$

Eigenvalues of $\boldsymbol{A}$

A square matrix $\boldsymbol{A}$ is said to be orthogonally diagonalizable if there exists an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{T}\boldsymbol{AP}$ is a diagonal matrix.

Definition (orthogonal diagonalization)

A square matrix $\boldsymbol{A}$ is said to be orthogonally diagonalizable if there exists an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

$$\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \lambda_1 & & & \mathbf{0} \\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_k \end{pmatrix}$$

Eigenvalues of $\boldsymbol{A}$

$\boldsymbol{P}$ is said to orthogonally diagonalize $\boldsymbol{A}$.

Theorem (characterization)

To check whether $\boldsymbol{A}$ is diagonalizable we need to determine if we have the required number (equal to the order of $\boldsymbol{A}$) of linearly independent eigenvectors.

How do we know whether a matrix is orthogonally diagonalizable?

A square matrix is orthogonally diagonalizable if and only if it is symmetric.

Proof: Omitted.

Algorithm (orthogonal diagonalization)

Given a symmetric matrix $\boldsymbol{A}$, we wish to find an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

Step 1: Solve the characteristic equation

as before

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$$

to find all the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$.

Step 2: For each eigenvalue λ_i,

as before

(a) Find a basis S_{λ_i} for the eigenspace E_{λ_i},

Algorithm (orthogonal diagonalization)

Step 2: For each eigenvalue λ_i,

(a) Find a basis S_{λ_i} for the eigenspace E_{λ_i},

NEW! (b) Apply Gram-Schmidt Process to transform S_{λ_i} into an orthonormal basis T_{λ_i}.

Step 3: Let $T = T_{\lambda_1} \cup T_{\lambda_2} \cup \ldots \cup T_{\lambda_k}$, say $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$,

put as columns

as before Then $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{pmatrix}$ is an orthogonal matrix that will orthogonally diagonalize $\boldsymbol{A}$.

Will this always work?

Step 3: Let $T = T_{\lambda_1} \cup T_{\lambda_2} \cup \ldots \cup T_{\lambda_k}$, say $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$,

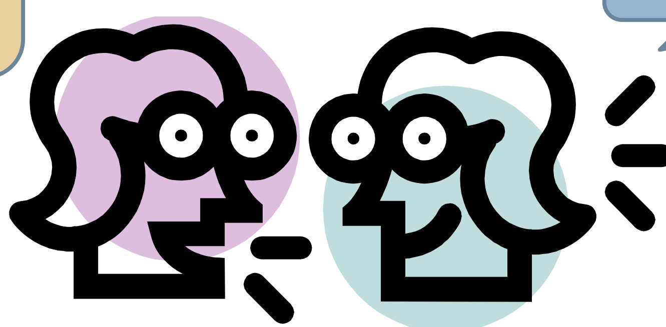
put as columns

as before

Then $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{pmatrix}$ is an orthogonal matrix that will orthogonally diagonalize $\boldsymbol{A}$.

Will I always have n vectors in the set T?

I promise you...



Some remarks

1) All the eignevalues λ_i are real numbers, that is,

the eigenvalues of symmetric matrices are all real.

2) From previous discussion, we know that if we factorise

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2}\ldots(\lambda - \lambda_k)^{r_k}$$

then for each $i = 1, \ldots, k,$

$$\boxed{r_1 + r_2 + \ldots + r_k = n}$$

no. of vectors in basis $S_{\lambda_i} = \dim(E_{\lambda_i}) \le r_i$

But for a symmetric matrix $\boldsymbol{A}$,

no. of vectors in basis $S_{\lambda_i} = \dim(E_{\lambda_i}) = r_i$ for all i.

Some remarks

2) From previous discussion, we know that if we factorise

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \ldots (\lambda - \lambda_k)^{r_k}$$

$\boxed{r_1 + r_2 + \ldots + r_k = n}$

But for a symmetric matrix $\boldsymbol{A}$,

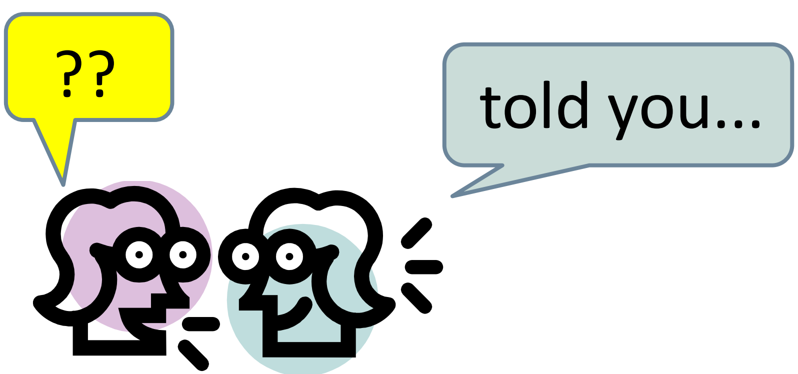
$$\boxed{\text{no. of vectors in basis } S_{\lambda_i} = \dim(E_{\lambda_i}) \boxed{= r_i \text{ for all } i.}}$$

$$\Rightarrow \text{no. of vectors in basis } T_{\lambda_i} = \dim(E_{\lambda_i}) = r_i \text{ for all } i.$$

So the number of vectors in $T (= T_{\lambda_1} \cup T_{\lambda_2} \cup \ldots \cup T_{\lambda_k})$

will always be

equal to $r_1 + r_2 + \ldots + r_k = n$.



Some remarks

3) In Step 3, the set

$$T = T_{\lambda_1} \cup T_{\lambda_2} \cup \ldots \cup T_{\lambda_k}, \text{ say } T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\},$$

is always an orthonormal set.

Note that each T_{λ_i} is an orthonormal set (since we performed Gram-Schmidt Process to transform S_{λ_i} to T_{λ_i}).

$$\{\underbrace{\quad\quad}_{}\}^{T_{\lambda_1}} \cup \{\quad\quad\}^{T_{\lambda_2}} \cup \ldots \cup \{\quad\quad\}^{T_{\lambda_k}}$$

all unit vectors

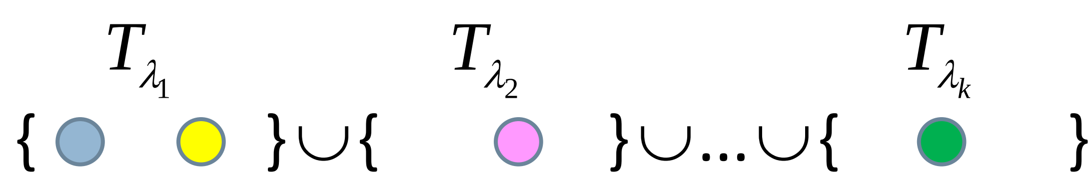
Some remarks

3) In Step 3, the set

$$T = T_{\lambda_1} \cup T_{\lambda_2} \cup \ldots \cup T_{\lambda_k}, \text{ say } T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\},$$

is always an orthonormal set.

Note that each T_{λ_i} is an orthonormal set (since we performed Gram-Schmidt Process to transform S_{λ_i} to T_{λ_i}).



$\cdot \quad = 0$ (due to Gram-Schmidt Process) $\qquad \cdot \quad = 0?$

(See Question 6.26)

Some remarks

3) In Step 3, the set

$$T = T_{\lambda_1} \cup T_{\lambda_2} \cup \ldots \cup T_{\lambda_k} \text{, say } T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\},$$

is always an orthonormal set.

4) Since T is always an orthonormal set with n vectors, the matrix $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{pmatrix}$ is always an orthogonal matrix.

Example

Find an orthogonal matrix $\boldsymbol{P}$ that orthogonally diagonalizes the following symmetric matrix.

$$\begin{pmatrix} 0 & -2 & 1 \\ -2 & 3 & -2 \\ 1 & -2 & 0 \end{pmatrix}$$

END OF LECTURE 19

Lecture 20

Linear transformations from R^n to R^m

(till end of Section 7.1)