Linear Algebra

Chapter 5: Orthogonality

University of Seoul
School of Computer Science
Minho Kim

Table of contents

Outline

More on Projection onto a Line

The standard matrix of a projection onto the line through the origin with direction vector $\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$: $P = \dfrac{1}{d_1^2 + d_2^2} \begin{bmatrix} d_1^2 & d_1 d_2 \\ d_1 d_2 & d_2^2 \end{bmatrix}$

- Can be written as

$$\begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix} = R_\theta \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} R_{-\theta} = R_{\theta-\frac{\pi}{2}} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} R_{\frac{\pi}{2}-\theta}$$

  cf) Excercise 26 of Sec 3.6 (p.230/222)

- $P = \mathbf{u}\mathbf{u}^T$ where $\mathbf{u} := \mathbf{d}/\|\mathbf{d}\|$. Why? (Geometrically)
  $\leftarrow P\mathbf{x} = \mathbf{u}(\mathbf{u}\cdot\mathbf{x}) = \mathbf{u}(\mathbf{u}^T\mathbf{x}) = (\mathbf{u}\mathbf{u}^T)\mathbf{x}$
- $P^T = P$ (symmetric)
  $\leftarrow P^T = (\mathbf{u}\mathbf{u}^T)^T = \mathbf{u}\mathbf{u}^T = P$
- $P^2 = P$ (idempotent)
  $\leftarrow P^2 = (\mathbf{u}\mathbf{u}^T)(\mathbf{u}\mathbf{u}^T) = \mathbf{u}(\mathbf{u}^T\mathbf{u})\mathbf{u}^T = \mathbf{u}(\mathbf{u}\cdot\mathbf{u})\mathbf{u}^T = \mathbf{u}\mathbf{u}^T = P$
- For a projection matrix $P \in \mathbb{R}^{2\times 2}$, the line onto which it projects vectors is the column space of P.
  $(\mathrm{col}(P) = \mathrm{span}\,(\mathbf{u})) \leftarrow P\mathbf{x} = (\mathbf{u}\mathbf{u}^T)\mathbf{x} = \mathbf{u}(\mathbf{u}^T\mathbf{x}) = (\mathbf{u}\cdot\mathbf{x})\mathbf{u}$

More on Projections onto Planes

Let

- $\mathscr{P}$ be a plane through the origin with unit normal $\mathbf{n}$ and
- $\mathrm{proj}_{\mathscr{P}}(\mathbf{v})$ be the projection of $\mathbf{v}$ onto $\mathscr{P}$.

Then

- What is $\mathrm{proj}_{\mathscr{P}}(\mathbf{v})$? $\Leftrightarrow$ Find c such that $(\mathbf{v} - c\mathbf{n}) \cdot \mathbf{n} = 0$.
  $\rightarrow c = \mathbf{v} \cdot \mathbf{n} \rightarrow \mathrm{proj}_{\mathscr{P}}(\mathbf{v}) = \mathbf{v} - (\mathbf{v} \cdot \mathbf{n})\mathbf{n}$
- Find $\mathrm{proj}_{\mathscr{P}}(\mathbf{v})$ using two direction vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ such that $\|\mathbf{u}_1\| = \|\mathbf{u}_2\| = 1$ and $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$ (orthogonal unit vectors)
  $\rightarrow P = \mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T$.
- $(\mathbf{v} - P\mathbf{v}) \perp \mathbf{u}_1$ and $(\mathbf{v} - P\mathbf{v}) \perp \mathbf{u}_2$?

$$\begin{aligned}
(\mathbf{v} - (\mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T)\mathbf{v}) \cdot \mathbf{u}_1 &= ((I - \mathbf{u}_1\mathbf{u}_1^T - \mathbf{u}_2\mathbf{u}_2^T)\mathbf{v})^T\mathbf{u}_1 \\
&= \mathbf{v}^T(I - \mathbf{u}_1\mathbf{u}_1^T - \mathbf{u}_2\mathbf{u}_2^T)\mathbf{u}_1 \\
&= \mathbf{v}^T(\mathbf{u}_1 - \mathbf{u}_1\mathbf{u}_1^T\mathbf{u}_1 - \mathbf{u}_2\mathbf{u}_2^T\mathbf{u}_1) \\
&= \mathbf{v}^T(\mathbf{u}_1 - \mathbf{u}_1 - 0) = 0
\end{aligned}$$

More on Projections onto Planes (cont'd)

- $P = P^T$ and $P^2 = P$. (symmetric and idempotent)
- $P = AA^T$ for some $A \in \mathbb{R}^{3\times 2}$.
  With $A := \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$,
  $$\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}^T = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{bmatrix} = \mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T$$
- $\mathrm{rank}(P) = 2$
  (See Theorem 3.25 (p.202) and Theorem 3.28, (p.205))
  $$\mathrm{rank}(AA^T) = \mathrm{rank}(A^T) = \mathrm{rank}(A) = \mathrm{rank}(A^TA)$$
  and
  $$A^TA = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 \end{bmatrix} = I_2$$

Outline

Orthogonal Sets of Vectors

What makes the standard basis of $\mathbb{R}^n$ good?

- orthogonal to each other
- unit length

Definition: Orthogonal Set

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is called an **orthogonal set** if all pairs of distinct vectors in the set are orthogonal – that is, if

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \quad \text{whenever } i \neq j \quad \text{for } i, j = 1, 2, \ldots, k$$

- Geometrically, they are mutually **perpendicular**.

Orthogonal Basis

Why is it good that the vectors are orthogonal?

Theorem 5.1

If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then these vectors are linearly independent.

$\rightarrow$ Can be used as a basis.

Definition

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthogonal set.

- Given two orthogonal vectors, how can we get the third orthogonal vector in $\mathbb{R}^3$?

  $\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2$

Orthogonal Basis (cont'd)

- How to compute the coordinate of a vector w.r.t. an orthogonal basis?

Theorem 5.2

Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then the unique scalars $c_1, \ldots, c_k$ such that

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$

are given by

$$c_i = \frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \quad \text{for } i = 1, \ldots, k$$

- $\mathbf{w} = \sum_{i=1}^{k} \left(\frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \right) \mathbf{v}_i = \sum_{i=1}^{k} \mathrm{proj}_{\mathbf{v}_i}(\mathbf{w})$

Orthonormal Basis

- Even better basis?

Definition: Orthonormal Set and Basis

A set of vectors in $\mathbb{R}^n$ is an **orthonormal set** if it is an orthogonal set of unit vectors. An **orthonormal basis** for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthonormal set.

- If $\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$ is an orthonormal set of vectors,

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

- How can we convert an orthogonal basis into an orthonormal basis? $\rightarrow$ Normalize vectors ($\mathbf{v}_i \rightarrow \mathbf{v}_i / \|\mathbf{v}_i\|$)

Orthonormal Basis (cont'd)

- How to compute the coordinate of a vector w.r.t. an orthonormal basis?

Theorem 5.3

Let $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k\}$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then

$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{w} \cdot \mathbf{q}_k)\mathbf{q}_k$$

and this representation is unique.

- Standard basis case ($\mathbf{q}_i = \mathbf{e}_i$)?

Orthogonal Matrices

- Given an orthonormal basis $\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$,

$$\begin{bmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_k\end{bmatrix}^T \begin{bmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_k\end{bmatrix} = ?$$

Theorem 5.4

The columns of an $m \times n$ matrix Q form an orthonormal set if and only if $Q^T Q = I_n$.

Definition: Orthogonal Matrix

An $n \times n$ matrix Q whose columns form an orthonormal set is called an **orthogonal matrix**.

- Square matrix
- Not an “orthonormal matrix”

Properties Orthogonal Matrices

Theorem 5.5

A square matrix Q is orthogonal if and only if $Q^{-1} = Q^T$.

- Example 5.7
  $\rightarrow$ Permutation matrices, rotation matrices

Theorem 5.6

Let Q be an $n \times n$ matrix. The following statements are equivalent:

a. Q is orthogonal.

b. $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ for every $\mathbf{x}$ in $\mathbb{R}^n$. (**isometry**: length-preserving)

c. $(Q\mathbf{x}) \cdot (Q\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for every $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. (angle-preserving)

Properties of Orthogonal Matrices (cont'd)

- $Q^T Q = I \Rightarrow QQ^T = I$.
  What does it mean?

Theorem 5.7

If Q is an orthogonal matrix, then its rows form an orthonormal set.

- More properties...

Theorem 5.8

Let Q be an orthogonal matrix.

a. Q^{-1} is orthogonal.

b. $\det Q = \pm 1$.

c. If λ is an eigenvalue of Q, then $|\lambda| = 1$.

d. If Q_1 and Q_2 are orthogonal $n \times n$ matrices, then so is $Q_1 Q_2$.

Outline

Orthogonal Complements

- How can we generalize the notion of "a normal vector to a plane" to higher dimensions?

Definition: Orthogonal Complement

Let W be a subspace of $\mathbb{R}^n$. We say that a vector $\mathbf{v}$ in $\mathbb{R}^n$ is **orthogonal to** W if $\mathbf{v}$ is orthogonal to every vector in W. The set of all vectors that are orthogonal to W is called the **orthogonal complement of** W, denoted $W^\perp$ ("W perp"). That is,

$$W^\perp = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \cdot \mathbf{w} = 0 \quad \text{for all } \mathbf{w} \in W\}$$

Example

- If $\mathbf{u}$ and $\mathbf{v}$ are the direction vectors of a plane in $\mathbb{R}^3$ and $\mathbf{n}$ is its normal vector,
  - $W = \mathrm{span}\,(\mathbf{u}, \mathbf{v})$
  - $W^\perp = \mathrm{span}\,(\mathbf{n})$

Properties of Orthogonal Complements

Theorem 5.9

Let W be a subspace in $\mathbb{R}^n$.

a. $W^\perp$ is a subspace of $\mathbb{R}^n$.

b. $(W^\perp)^\perp = W$

c. $W \cap W^\perp = \{\mathbf{0}\}$

d. If $W = \mathrm{span}\,(\mathbf{w}_1, \ldots, \mathbf{w}_k)$, then $\mathbf{v}$ is in $W^\perp$ if and only if $\mathbf{v} \cdot \mathbf{w}_i = 0$ for all $i = 1, \ldots, k$.

- Is there a nonzero vector $\mathbf{x}$ such that
  - $\mathbf{x} \in W$ and $\mathbf{x} \in W^\perp$?
  - $\mathbf{x} \perp W$ and $\mathbf{x} \perp W^\perp$?
- $W \cup W^\perp =$?
- $\dim W + \dim(W^\perp) =$? (Theorem 5.13)

Fundamental Subspaces of a Matrix

- Orthogonal complements and the subspaces associated with an $m \times n$ matrix.

Theorem 5.10

Let A be an $m \times n$ matrix. Then the orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(\mathrm{row}(A))^{\perp} = \mathrm{null}(A) \quad \text{and} \quad (\mathrm{col}(A))^{\perp} = \mathrm{null}(A^T)$$

- $\mathrm{null}(A^T)$ is called *left nullspace* of A.
- With $A \in \mathbb{R}^{m \times n}$,
  - For $\mathbf{x}_r \in \mathrm{row}(A)$, $A\mathbf{x}_r \in \mathrm{col}(A)$.
    $\mathrm{row}(A) \xrightarrow{T_A} \mathrm{col}(A) \subset \mathbb{R}^m$
  - If $\mathbf{x}_n \in \mathrm{null}(A)$, then $A\mathbf{x}_n = \mathbf{0}$.
    $\mathrm{null}(A) \xrightarrow{T_A} \{\mathbf{0}\} \subset \mathbb{R}^m$
  - See Figure 5.6 on p.377.

Orthogonal Projections

- How can we generalize "the projection of a vector onto a line or a plane"?

Definition: Orthogonal Projection

Let W be a subspace of $\mathbb{R}^n$ and let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal basis for W. For any vector $\mathbf{v}$ in $\mathbb{R}^n$, the **orthogonal projection of $\mathbf{v}$ onto W** is defined as

$$\begin{aligned}\operatorname{proj}_W(\mathbf{v}) &= \left(\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)\mathbf{u}_1 + \cdots + \left(\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\right)\mathbf{u}_k \\ &= \operatorname{proj}_{\mathbf{u}_1}(\mathbf{v}) + \cdots + \operatorname{proj}_{\mathbf{u}_k}(\mathbf{v})\end{aligned}$$

The **component of $\mathbf{v}$ orthogonal to W** is the vector $\operatorname{perp}_W(\mathbf{v}) = \mathbf{v} - \operatorname{proj}_W(\mathbf{v})$

- "Orthogonal decomposition"
- See Figure 5.8 on p.394/380.

Orthogonal Decomposition

- Is the orthogonal decomposition unique?

Theorem 5.11: The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Then there are unique vectors $\mathbf{w}$ in W and $\mathbf{w}^\perp$ in $W^\perp$ such that

$$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$$

- $\mathrm{proj}_W(\mathbf{v})$ and $\mathrm{perp}_W(\mathbf{v})$ do not depend on the choice of orthogonal basis.
- Can be used to prove

$$(W^\perp)^\perp = W$$

- Orthogonal decomposition & fundamental subspaces of $A \in \mathbb{R}^{m\times n}$
  1. For any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$ ($\mathbf{x}_r \in \mathrm{row}(A)$ & $\mathbf{x}_n \in \mathrm{null}(A)$)
  2. $A\mathbf{x} = A(\mathbf{x}_r + \mathbf{x}_n) = A\mathbf{x}_r + \mathbf{0} = A\mathbf{x}_r \in \mathrm{col}(A)$

Orthogonal Decomposition (cont'd)

- Relationship between the dimension of W and $W^\perp$

Theorem 5.13

If W is a subspace of $\mathbb{R}^n$, then

$$\dim W + \dim W^\perp = n$$

- Special case:

Corollary 5.14: The Rank Theorem

If A is an $m \times n$ matrix, then

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n$$

- $\operatorname{rank}(A) + \operatorname{nullity}(A^T) = m$

3F From the row space to the column space, A is actually invertible. Every vector b in the column space comes from exactly one vector x_r in the row space.

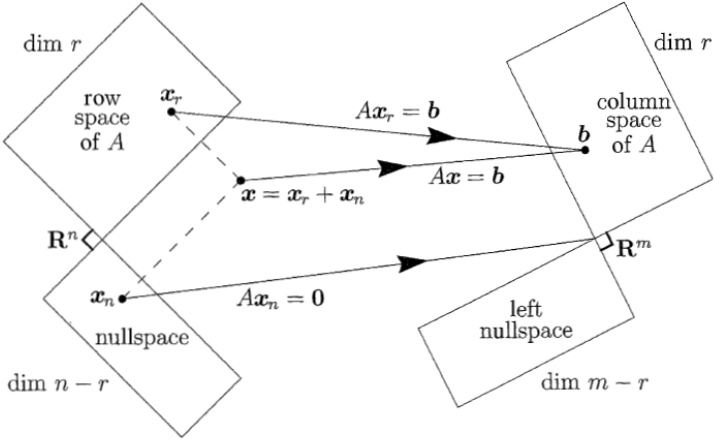

Figure 3.4 The true action $Ax = A(x_{\text{row}} + x_{\text{null}})$ of any m by n matrix.

* image courtesy of Gilbert Strang

Outline

The Gram-Schmidt Process

- Given a subspace, how can we construct an orthogonal/orthonormal basis?
- The Gram-Schmidt Process: Starting from an arbitrary basis for a subspace, "orthogonalize" it one vector at a time.
  $\rightarrow$ Example 5.12

The Gram-Schmidt Process (cont'd)

Theorem 5.15: The Gram-Schmidt Process

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ be a basis for a subspace W of $\mathbb{R}^n$ and define the following:

$$
\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1, & W_1 &= \operatorname{span}(\mathbf{x}_1) \\
\mathbf{v}_2 &= \mathbf{x}_2 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1, & W_2 &= \operatorname{span}(\mathbf{x}_1, \mathbf{x}_2) \\
\mathbf{v}_3 &= \mathbf{x}_3 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{x}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2, & W_3 &= \operatorname{span}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \\
&\vdots \\
\mathbf{v}_k &= \mathbf{x}_k - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_k}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{x}_k}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2 - \cdots \\
&\quad - \left(\frac{\mathbf{v}_{k-1} \cdot \mathbf{x}_k}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}}\right)\mathbf{v}_{k-1}, & W_k &= \operatorname{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)
\end{aligned}
$$

Then for each $i = 1, \ldots, k$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is an orthogonal basis for W_i. In particular, $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for W.

The QR Factorization

- Factorization of a matrix according to the Gram-Schmidt process.
- Applications:
  Approximation of eigenvalues (p.409/395), least squares approximation (Chap. 7)

$$W_i = \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_i) = \operatorname{span}(\mathbf{q}_1, \ldots, \mathbf{q}_i)$$

$$\rightarrow \mathbf{a}_i = r_{1i}\mathbf{q}_1 + r_{2i}\mathbf{q}_2 + \cdots + r_{ii}\mathbf{q}_i, \quad \text{for } i = 1, \ldots, n$$

$$\rightarrow A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix} = QR$$

The QR Factorization (cont'd)

Theorem 5.16: The QR Factorization

Let A be an $m \times n$ matrix with linearly independent columns. Then A can be factored as $A = QR$, where Q is an $m \times n$ matrix with orthonormal columns and R is an invertible upper triangular matrix.

- Why is R invertible?
- How can we find R? (Example 5.15)
  $\rightarrow R = Q^T A$

Outline

Real Symmetric Matrices

- Does a square matrix with real entries have real eigenvalues?
  $\rightarrow$ No. $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$
- Are all square matrices diagonalizable?
  $\rightarrow$ No. (Example 4.25 on p.315/301)

Real symmetric matrices are good!

- All eigenvalues are real.
- Always diagonalizable.

Real Symmetric Matrices (cont'd)

Definition: Orthogonally Diagonalizable

A square matrix A is **orthogonally diagonalizable** if there exists an orthogonal matrix Q and a diagonal matrix D such that $Q^TAQ = D$.

- Example 5.16
- Why is it good to be orthogonally diagonalizable?

Theorem 5.17

If A is orthogonally diagonalizable, then A is symmetric.

- How about the converse? Is every symmetric matrix is orthogonally diagonalizable?
  $\rightarrow$ Theorem 5.20 (p.414/400)

The (Real) Spectral Theorem

Theorem 5.18

If A is a real symmetric matrix, then the eigenvalues of A are real.

- Theorem 4.20 (p.308/294): “Eigenvectors corresponding to distinct eigenvalues are linearly independent.”
  $\rightarrow$ How about symmetric matrices?

Theorem 5.19

If A is a symmetric matrix, then any two eigenvectors corresponding to distinct eigenvalues of A are orthogonal.

Theorem 5.20: The Spectral Theorem

Let A be an $n \times n$ real matrix. Then A is symmetric **if and only if** it is orthogonally diagonalizable.

Spectral Decomposition

$$
\begin{aligned}
A = QDQ^T &= \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 \mathbf{q}_1 & \cdots & \lambda_n \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \\
&= \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T
\end{aligned}
$$

- “Projection form of the Spectral Theorem”
- Steps
    1. $A = PDP^{-1}$ (diagonalization)
    2. $P \to Q$ (Gram-Schmidt process)
    3. $A = QDQ^T$ (orthogonal diagonalization)

Matrix Approximation of Symmetric Matrices

- How much storage required to store an $n \times n$ symmetric matrix A? $\to$ $n(n+1)/2$

1. Orthogonally diagonalize A with eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$:

$$A = QDQ^T$$

2. Expand it and take only first k ($k < n/2$) terms:

$$B := \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \cdots + \lambda_k \mathbf{q}_k \mathbf{q}_k^T$$

3. Now
   - B requires only $k(n+1)$ numbers to be stored. (Why?)
   - the least-square error between A and B,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} - b_{ij})^2$$

   is

$$\sum_{i=k+1}^{n} \lambda_i^2.$$

- More in §7.4 The Singular Value Decomposition