Announcements

• HW 14 serves as review. It contains roughly one problem from each textbook section that we discussed. Most problems are focused on understanding concepts rather than computations.

• HW 14 will close Monday April 22, as usual.

• The Final Exam will take place Thursday, April 25, 8:00-10:00.

• it will cover the whole semester about evenly.

• Today: Study Session, 11:50, JTB 120

• Today and tomorrow: review of chapters 5 and 6

• Wednesday: Exam 4 on chapter 5 and 6. It will have a total of 6 questions, 3 on chapter 5, and 3 on chapter 6.
Reviews of Chapters 5 and 6

Unless stated otherwise, throughout this review, \( A \) is a square, \( n \times n \) real matrix.

6. Eigenvalues and Eigenvectors

- An eigenvector of a square \((n \times n)\) matrix \( A \) is a non-zero vector \( x \) such that

\[
Ax = \lambda x
\]

for some scalar \( \lambda \). \( \lambda \) is called the eigenvalue of \( A \) corresponding to the eigenvector \( x \). \( x \) is an eigenvector corresponding to the eigenvalue \( \lambda \).

- The pair \((\lambda, x)\) is sometimes called an eigenpair of \( A \).

\[\text{Note that any non-zero scalar multiple of an eigenvector is also an eigenvector, with the same eigenvalue.}\]

\[\text{the main difference between linear systems and eigenvalue problems is that eigenvalue problems are nonlinear!}\]

- More insight can be gained by writing

\[
Ax = \lambda x
\]

as

\[
Ax - \lambda x = (A - \lambda I)x = 0.
\]
• Any eigenvector is a non-trivial solution of the homogeneous linear system

\[(A - \lambda I)x = 0.\]  \hspace{1cm} (4)

• Every eigenvector is in the nullspace of \(A - \lambda I\).

• Every non-zero vector in the nullspace of \(A - \lambda I\) is an eigenvector of \(A\).

• A square homogeneous linear system has a non-trivial solution if and only if the coefficient matrix is singular.

\[\text{thus } \lambda \text{ is an eigenvalue of } A \text{ if and only if } A - \lambda I \text{ is singular.}\]

\[\text{Upshot: we have one more characterization of singularity. A square matrix } A \text{ is singular if and only if } 0 \text{ is an eigenvalue of } A. \text{ It is invertible if and only if all eigenvalues of } A \text{ are non-zero.}\]

• Suppose \(x_i, i = 1, \ldots, m\) are eigenvectors corresponding to the same eigenvalue \(\lambda\). Then any (non-zero) linear combination of the eigenvectors is also an eigenvector:

\[A \sum_{i=1}^{m} \alpha_i x_i = \sum_{i=1}^{n} \alpha_i Ax_i = \sum_{i=1}^{n} \alpha_i \lambda x_i = \lambda \sum_{i=1}^{n} \alpha_i x_i.\]  \hspace{1cm} (5)

• Thus, if we add the zero vector to the set of eigenvectors corresponding to a specific eigenvalue, that set is a linear space, the nullspace
of $A - \lambda I$. That space is also called the **eigenspace of $A$ corresponding to** $\lambda$.

- **Important example:** The eigenvalues of a **triangular** matrix are the **diagonal entries**, because if $A$ is triangular and $\lambda$ is an eigenvalue then $A - \lambda I$ is a triangular matrix with at least one zero entry on the diagonal. It is thus singular.

=row operations do not preserve eigenvalues or eigenvectors!

- A matrix is singular if and only if its determinant is zero. Thus we get the key result:

$$\lambda \text{ is an e.v. } \iff \det(A - \lambda I) = 0. \quad (6)$$

- The equation

$$\det(A - \lambda I) = 0. \quad (7)$$

is the **characteristic equation** of $A$.

- The function $f(\lambda) = |A - \lambda I|$ is a polynomial of degree $n$ with leading coefficient $(-1)^n$.

- We can see this using a cofactor expansion or the formula

$$\det A = \sum_{\sigma} \text{sign}(\sigma) \prod_{i=1}^{n} a_{i\sigma_i} \quad (8)$$

where the sum goes over all $n!$ permutations of the set $\{1, 2, \ldots, n\}$ and the large symbol
π indicates the product of \( n \) factors, one from each row \( i \), and the column \( \sigma_i \).

- The polynomial

\[
p(\lambda) = \det(A - \lambda I)
\]  \hspace{1cm} (9)

is the **characteristic polynomial** of \( A \)

\[\begin{align*}
\text{The eigenvalues of } A \text{ are the roots of the characteristic polynomial.}
\end{align*}\]

- This means
  1. There are precisely \( n \) of them, properly counting multiplicity.
  2. They may be repeated.
  3. They may be complex.
  4. If there are complex eigenvalues then they occur in conjugate complex pairs.

- The natural way to compute eigenvalues and eigenvectors by hand proceeds in two steps:
  1. Compute the characteristic polynomial and find its roots.
  2. For each eigenvalue \( \lambda \) find the nullspace of \( A - \lambda I \).

- This works well only for small matrices with exactly known entries.

- However, the opposite process, computing roots of polynomials by computing the eigenvalues of a suitable matrix works very well.
• For every polynomial $p$ of degree $n$ with leading term $(-1)^n$ there exists a matrix $A$ whose characteristic polynomial is $p$. Check:

\[
\det \left( \begin{bmatrix} \alpha_{n-1} & \alpha_{n-2} & \cdots & \alpha_1 & \alpha_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} - \lambda I \right) = (-1)^n \left[ \lambda^n - \sum_{j=0}^{n-1} \alpha_j \lambda^j \right].
\]

(10)

• Central in eigenvalue calculations is the concept of similarity.

• **Definition:** Two matrices $A$ and $B$ are similar if there is a non-singular matrix $P$ such that

\[ B = P^{-1}AP. \]

(11)

• Similar matrices have the same eigenvalues, and their eigenvectors are related in a straightforward way. To see this suppose that

\[ Ax = \lambda x \]

(12)

and note that

\[ B(P^{-1}x) = P^{-1}APP^{-1}x = P^{-1}\lambda x = \lambda(P^{-1}x). \]

(13)

• In other words, the eigenvectors of $B$ are those of $A$, multiplied with $P^{-1}$. 

Math 2270-6 Notes of 4/15/19 page 6
Another way to see that similar matrices have the same eigenvalues is to observe that their characteristic polynomials are the same. Using the multiplicative property of determinants and the fact that the determinant of the inverse is the reciprocal of the determinant of the original matrix we see

\[ |B - \lambda I| = |P^{-1}AP - \lambda P^{-1}IP| \]
\[ = |P^{-1}(A - \lambda I)P| \]
\[ = |P^{-1}|A - \lambda I||P| \]
\[ = |A - \lambda I| \] \hspace{1cm} (14)

**Definition:** A matrix is **diagonalizable** if it is similar to a diagonal matrix.

In other words, \( A \) is diagonalizable if there exists a diagonal matrix \( D \) and a non-singular matrix \( P \) such that

\[ D = P^{-1}AP. \] \hspace{1cm} (15)

This equation can be rewritten as

\[ AP = PD. \] \hspace{1cm} (16)

Suppose

\[ P = [v_1 \quad v_2 \quad \ldots \quad v_n] \] \hspace{1cm} (17)
and

\[
D = \begin{bmatrix}
\lambda_1 & 0 & 0 & \ldots & 0 \\
0 & \lambda_2 & 0 & \ldots & 0 \\
0 & 0 & \lambda_3 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \lambda_n
\end{bmatrix}
\] (18)

- Note that the equation for the \(i\)-th column in (16) is precisely the eigenvector equation

\[
A\mathbf{v}_i = \lambda_i \mathbf{v}_i.
\] (19)

A matrix is diagonalizable if and only if it has \(n\) linearly independent eigenvectors. The similarity transform to diagonal form is the matrix of eigenvectors and the similar diagonal matrix has the eigenvalues along the diagonal.

- A matrix that is not diagonalizable is called **defective**.

- A matrix is not defective if and only if it has a set of \(n\) linearly independent eigenvectors.

Invertibility is unrelated to Diagonaliz-
ability.

\textbf{defective:} \[
\begin{bmatrix}
0 & 1 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
1 & 1 \\
0 & 1
\end{bmatrix}
\]

\textbf{diagonalizable:} \[
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]

(20)

- It is sometimes useful to be able to construct a matrix with given eigenvalues and eigenvectors. Note that

\[ D = P^{-1}AP \]  \hspace{1cm} (21)

is equivalent to

\[ A = PD P^{-1}. \]  \hspace{1cm} (22)

Suppose you want to construct a matrix \( A \) with given eigenvalues and given eigenvectors. Proceed as follows:

1. Collect the eigenvectors into the matrix \( P \) as before.
2. Compute \( P^{-1} \).
3. Compute \[ A = PD P^{-1}. \]  \hspace{1cm} (23)
• It is not always possible to diagonalize a matrix. However, for all matrices $A$ there exists a similarity transform to its **Jordan Canonical Form**$^{-1}$ (named after Camille Jordan, 1838-1922).

• The JCF is a block diagonal matrix

$$P^{-1}AP = \begin{bmatrix} J_1 & 0 & \ldots & 0 \\ 0 & J_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & J_k \end{bmatrix}$$

(24)

where each diagonal block is of the form

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 & 0 \\ 0 & 0 & \lambda_i & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_i & 1 \\ 0 & 0 & 0 & \ldots & 0 & \lambda_i \end{bmatrix}$$

(25)

• Apart from reordering the diagonal blocks the JCF is unique.

• Each Jordan block $J_i$ corresponds to one eigenvector with eigenvalue $\lambda_i$.

• A matrix is diagonalizable if and only if all of its Jordan blocks are $1 \times 1$.

$^{-1}$ The textbook mentions the Jordan Canonical Form in a footnote on page 294.
• The **algebraic multiplicity** of an eigenvalue is its order as a root of the characteristic polynomial. Its **geometric multiplicity** is the dimension of its eigenspace.

• Here is an example. Suppose the Jordan form of a matrix is given by

\[
J = \begin{bmatrix}
2 & 1 & & & & & & \\
& 2 & & & & & & \\
& & 3 & 1 & & & & \\
& & & 3 & 1 & & & \\
& & & & 3 & & & \\
& & & & & 4 & & \\
& & & & & & 4 & \\
& & & & & & & 5
\end{bmatrix}
\]

(26)

• Entries indicated by dots are zero.

• The characteristic polynomial of this matrix is

\[
p(\lambda) = |J - \lambda I| = (2 - \lambda)^3 (3 - \lambda)^4 (4 - \lambda)^2 (5 - \lambda).
\]

(27)

The number 2 is an eigenvalue of algebraic multiplicity 3 and geometric multiplicity 2, 3 is an eigenvalue of algebraic multiplicity 2 and geometric multiplicity 2, 4 is an eigenvalue of algebraic and geometric multiplicity 2, and 5 is an eigenvalue of algebraic and geometric multiplicity 1. The dimension of the
space spanned by all eigenvectors is the sum of the geometric multiplicities which is 7. The matrix is defective.

- A set of eigenvectors corresponding to distinct eigenvalues is linearly independent. (The word “distinct” means that no two of the eigenvalues are equal.)

- Recall that a matrix is diagonalizable if it has a set of $n$ linearly independent eigenvectors.

- Thus a matrix with distinct eigenvalues is diagonalizable.

- This implies, for example, that the JCF can be computed only in exact arithmetic.

- A non-diagonalizable matrix must have multiple eigenvalues.

The most important thing to know about complex eigenvalues is that symmetric real matrices don’t have any! The textbook addresses this issue in problem 24 on page 303 (and later in chapter 7).

- But the argument is quite simple.

- For any matrix $A$ or vector $x$ let

$$A^H = \overline{A}^T \quad \text{and} \quad x^H = \overline{x}^T \quad (28)$$

where the bar denotes conjugate complex.
• A complex matrix $A$ is **Hermitian** if

$$A = \bar{A}^T.$$  \hspace{1cm} (29)

• We will show that the eigenvalues of a Hermitian matrix are real.

Note that symmetric real matrices are special cases of Hermitian matrices.

• Suppose

$$Ax = \lambda x$$  \hspace{1cm} (30)

where $A = A^H$, and $A$, $\lambda$, and $x$ are all possibly complex. Taking the conjugate complex on both sides turns this into

$$x^H A^H = x^H A = \bar{\lambda} x^H.$$  \hspace{1cm} (31)

Left multiplying with $x^H$ in (30) and right multiplying with $x$ in (31) gives

$$x^H Ax = \lambda x^H x \quad \text{and} \quad x^H Ax = \bar{\lambda} x^H x.$$  \hspace{1cm} (32)

Thus

$$\lambda x^H x = \bar{\lambda} x^H x.$$  \hspace{1cm} (33)

This implies that $\lambda = \bar{\lambda}$, i.e., $\lambda$ is real.

• **Gershgorin Theorem.** Suppose $A$ is an $n \times n$ matrix, and $\lambda$ is one of its eigenvalues. Then, for some $i \in \{1, 2, \ldots, n\}$

$$|a_{ii} - \lambda| \leq \sum_{j \neq i} |a_{ij}|.$$  \hspace{1cm} (34)

\footnote{named after Charles Hermite, 1822–1901.}
• In other words, every eigenvalue lies in some circle whose center is a diagonal entry of $A$, and whose radius equals the sum of the absolute values of the off-diagonal entries in that row.

• Those circles are referred to as the Gershgorin Circles.

• To see this suppose $x$ is an eigenvector of the $n \times n$ matrix $A$, with corresponding eigenvalue $\lambda$. Thus

$$Ax = \lambda x. \quad (35)$$

• Since an eigenvector is determined only up to a non-zero factor we may assume that $x$ is normalized such that

$$\max_{j=1,\ldots,n} |x_j| = x_i = 1 \quad (36)$$

for some $i$ in $\{1, 2, \ldots, n\}$. This fixes $i$. If there are several such indices $i$ we pick any particular one of them.

• The $i$-th component of the vector equation (36) is

$$\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i = \lambda. \quad (37)$$

• Subtracting $a_{ii}x_i = a_{ii}$ on both sides gives the equation

$$\lambda - a_{ii} = \sum_{j \neq i} a_{ij}x_j \quad (38)$$
• Taking absolute values on both sides, applying the triangle inequality, and observing that $|x_j| \leq 1$ for all $j$ shows that $\lambda$ lies in the Gershgorin Circle centered at $x_i$:

$$|\lambda - a_{ii}| = \left| \sum_{j \neq i} a_{ij} x_j \right|$$

$$\leq \sum_{j \neq i} |a_{ij} x_j|$$

$$= \sum_{j \neq i} |a_{ij}| |x_j|$$

$$\leq \sum_{j \neq i} |a_{ij}|$$

(39)

It’s not true in general that every Gershgorin Circle contains an eigenvalue.

On the other hand, it is true that any union of $k$ Gershgorin Circles that does not overlap with any of the remaining Gershgorin Circles contains precisely $k$ eigenvalues, counting multiplicity.

6. Orthogonality and Least Squares

• The inner product, previously called the dot product, of two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,
is defined to be

\[
\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \sum_{i=1}^{n} u_i v_i.
\]

\((\mathbf{u}, \mathbf{v})\)

\langle \mathbf{u}, \mathbf{v} \rangle

- **Theorem 1, p. 333.** Let \(\mathbf{u}, \mathbf{v}\) and \(\mathbf{w}\) be vectors in \(\mathbb{R}^n\), and \(c\) be a scalar. Then
  
  a. \(\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}\)
  
  b. \((\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}\)
  
  c. \((c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})\)
  
  d. \(\mathbf{u} \cdot \mathbf{u} \geq 0, \quad \text{and} \quad \mathbf{u} \cdot \mathbf{u} = 0 \implies \mathbf{u} = \mathbf{0}.\)

- The **length** or **norm** of a vector \(\mathbf{v}\) is defined by
  
  \[\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.\]  

  \(41\)

- **Definition:** Two vectors \(\mathbf{u}\) and \(\mathbf{v}\) are **orthogonal** (or **perpendicular**) if
  
  \(\mathbf{u} \cdot \mathbf{v} = 0.\)  

  \(42\)

- the zero vector is orthogonal to all vectors in \(\mathbb{R}^n\).

- Also called **Standard Norm**, **Euclidean Norm**, or **2-norm**.
Suppose $W$ is a subspace of $\mathbb{R}^n$. Then the set

$$W^\perp = \{ z : z \text{ is orthogonal to all vectors in } W \}$$

(43)

is a linear space, called the orthogonal complement of $W$.

$W^\perp$ is read as "$W$-perpendicular" or, more commonly, just "$W$-perp".

Example: line and plane in $\mathbb{R}^3$.

**Theorem 3**, p. 337: Let $A$ be an $m \times n$ matrix. 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 $A^T$

$$(\text{Row} A)^\perp = \text{Nul} \quad \text{and} \quad (\text{Col} A)^\perp = \text{Nul} A^T.$$ 

(44)

A set of vectors $\{u_1, u_2, \ldots, u_p\}$ from $\mathbb{R}^n$ is an orthogonal set if each pair of distinct vectors from that set is orthogonal, i.e.,

$$i \neq j \implies u_i \cdot u_j = 0.$$ 

(45)

**Theorem 4**, p. 340, textbook. If

$$S = \{u_1, u_2, \ldots, u_p\}$$

(46)

is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then $S$ is linearly independent. (Hence $S$ is a basis of span($S$).)
• Naturally, an **orthogonal basis** for a subspace $W$ of $\mathbb{R}^n$ is a basis for $W$ that is also an orthogonal set.

• Orthogonal Bases are nice! You can compute coefficients without solving a linear system.

• Suppose

$$B = \{u_1, u_2, \ldots, u_p\}$$

is a basis of a subspace $W$ of $\mathbb{R}^n$,\n
$$B = [u_1, u_2, \ldots, u_p], \quad (48)$$

and $y$ is a vector in $W$. Then, in general, computing the coordinate vector

$$[y]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}$$

of $y$ requires the solution of the linear system

$$B[y]_B = y. \quad (50)$$

• However, if $B$ is an orthogonal basis we can compute the components of $[y]_B$ directly:

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j}. \quad (51)$$
• **Theorem 6**, p. 345. An $m \times n$ matrix $U$ has orthonormal columns if and only if

$$U^TU = I$$

(52)

(where $I$ is the $n \times n$ identity matrix.).

• **Theorem 7**, p. 345. Let $U$ be an $m \times n$ matrix with orthonormal columns, and let $x$ and $y$ be vectors in $\mathbb{R}^n$. Then:
  
  a. $\|Ux\| = \|x\|
  
  b. $(Ux) \cdot (Uy) = x \cdot y$
  
  c. $(Ux) \cdot (Uy) = 0$ if and only if $x \cdot y = 0$

• The Pythagorean Theorem states that

$$\|u \pm v\|^2 = \|u\|^2 + \|v\|^2 \iff u \cdot v = 0.$$  

(53)

• The **orthogonal projection** of a vector $v$ onto a vector $u$ is given by

$$\text{proj}_u v = \frac{u \cdot v}{uu} u.$$  

(54)

• **Theorem 8.** Let $W$ be a subspace of $\mathbb{R}^n$. Then each $y$ in $\mathbb{R}^n$ can be written uniquely in the form

$$y = \hat{y} + z$$

(55)

where $\hat{y}$ is in $W$ and $z$ is in $W^\perp$.

• This is the **orthogonal Decomposition theorem.** The vector $\hat{y}$ in (55) is called the **orthogonal projection of $y$ onto $W$.**
• The textbook uses the notation

\[ \hat{y} = \text{proj}_W y. \] \hfill (56)

• **Best Approximation Theorem** (Theorem 9, p. 352) Let \( W \) be a subspace of \( \mathbb{R}^n \), let \( y \) be any vector in \( \mathbb{R}^n \), and let \( \hat{y} \) be the orthogonal projection of \( y \) onto \( W \). Then \( \hat{y} \) is the closest point \( W \) to \( y \), in the sense that

\[ \| y - v \| < \| y - \hat{y} \| \] \hfill (57)

for all \( v \) in \( W \) distinct from \( \hat{y} \).

• **Theorem 10**, p. 353. If \( \{u_1, u_2, \ldots, u_p\} \) is an orthonormal basis for a subspace \( W \) of \( \mathbb{R}^n \), then

\[ \hat{y} = \text{proj}_W y = \sum_{i=1}^{p} (y \cdot u_i) u_i. \] \hfill (58)

If \( U = [u_1 \quad u_2 \ldots u_p] \), then

\[ \text{proj}_W y = UU^T y \] \hfill (59)

for all \( y \) in \( \mathbb{R}^n \).

• We considered three versions of the Gram-Schmidt Process.

• Version 1: is described by **Theorem 11**, page 357, textbook: Given a basis

\[ \{x_1, x_2, \ldots, x_p\} \] \hfill (60)
for a non-zero subspace $W$ of $\mathbb{R}^n$, define

$$v_1 = x_1$$

$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$

$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2$$

$$\vdots$$

$$v_p = x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - \ldots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}$$

(61)

Then $\{v_1, \ldots, v_p\}$ is an orthogonal basis for $W$. In addition,

$$\text{span} \{v_1, \ldots, v_k\} = \text{span} \{x_1, \ldots, x_k\} \quad \text{for} \quad k = 1, 2, \ldots, p.$$  

(62)

- Version 2 is just a more compact notation for the process. For $k = 1, \ldots, p$ define

$$v_k = x_k - \sum_{i=1}^{k-1} \frac{x_k \cdot v_i}{v_i \cdot v_i} v_i.$$  

(63)

- Version 3 combines normalization with orthogonalization: For $k = 1, \ldots, p$ define

$$\begin{cases}
  w_k &= x_k - \sum_{i=1}^{k-1} (x_k \cdot v_i) v_i \\
  v_k &= \frac{w_k}{\|w_k\|}
\end{cases}$$  

(64)
• Definition: A square matrix $Q$ is **orthogonal** if its columns form an **orthonormal** set.

• This means that

$$Q^T Q = I,$$  \hspace{1cm} (65)

i.e., $Q$ is invertible, and

$$Q^{-1} = Q^T.$$  \hspace{1cm} (66)

(see textbook, page 346.)

• **Theorem 12**, page 359, textbook. If $A$ is an $m \times n$ matrix with linearly independent columns, then $A$ can be factored as

$$A = QR$$  \hspace{1cm} (67)

where $Q$ is an $m \times n$ matrix whose columns form an orthonormal basis for $\text{Col}(A)$ and $R$ is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

• Suppose we have an overdetermined linear system

$$Ax = b$$  \hspace{1cm} (68)

• Here $A$ is $m \times n$, $x$ is in $\mathbb{R}^n$, $b$ is in $\mathbb{R}^m$, and $m \geq n$ (and typically, $m > n$).

• Usually, the system (68) will not have a solution. In that case, the next best thing is to solve the alternative problem

$$\|Ax - b\| = \min$$  \hspace{1cm} (69)
• In other words (the words of our textbook), we want to find a vector \( \hat{x} \) in \( \mathbb{R}^n \) such that

\[
\|b - A\hat{x}\| \leq \|b - Ax\| \quad (70)
\]

for all \( x \) in \( \mathbb{R}^n \).

• The textbook calls such an \( \hat{x} \) a **Least Squares Solution** of

\[
Ax = b. \quad (71)
\]

• I would call it a solution of

\[
\|Ax - b\| = \text{min}. \quad (72)
\]

• First: **Theorem 13** (p. 363) The set of least square solutions of \( Ax = b \) coincides with the nonempty set of solutions of the **normal equations**

\[
A^T Ax = A^T b. \quad (73)
\]

• **Theorem 14** (p. 365) Let \( A \) be an \( m \times n \) matrix. The following statements are logically equivalent. (This means they are either all true or all false):

  a. The equation \( Ax = b \) has a unique least squares solution for each \( b \) in \( \mathbb{R}^m \).

  b. The columns of \( A \) are linearly independent.

  c. The matrix \( A^T A \) is invertible.
\( \mathbb{R}^m \)

\[ a_i \cdot (b - Ax) = 0 \quad A = [a_1, \ldots, a_n] \]

\[ A^T (b - Ax) = 0 \]

\[ A^T A x = A^T b \]
• Suppose we write

\[
A = QR
\]  

(74)

where

\[
Q = \begin{pmatrix}
    n & m - n \\
    Q_1 & Q_2 \\
\end{pmatrix}
\]  

(75)

is orthogonal and

\[
R = \begin{pmatrix}
    n \\
    m - n \\
    R_1 \\
    0
\end{pmatrix}
\]  

(76)

with \( R_1 \) being upper triangular.

• Earlier we discussed how to obtain

\[
A = Q_1 R_1,
\]  

(77)

for example by the Gram-Schmidt Process.

• To get \( Q \) from \( Q_1 \) we simply add vectors to the orthonormal basis of the column space of \( A \) to get an orthonormal basis of \( \mathbb{R}^m \).

• We won’t actually need \( Q_2 \), but it’s useful to describe the idea.

• A significant property of an orthogonal matrix is that multiplying with it does not alter the norm of a vector:

\[
\|Qx\|^2 = (Qx)^T(Qx) = x^TQ^TQx = x^Tx = \|x\|^2.
\]  

(78)
• Using

\[ A = QR \quad \text{and} \quad Q^T A = R \quad (79) \]

we obtain

\[
\|Ax - b\|^2 = \|Q^T(Ax - b)\|^2 \\
= \|Q^T Ax - Q^T b\|^2 \\
= \left\| \begin{pmatrix} R_1 x \\ 0 \end{pmatrix} - \begin{pmatrix} Q^T_1 b \\ Q^T_2 b \end{pmatrix} \right\|^2 \\
= \|R_1 x - Q^T_1 b\|^2 + \|Q^T_2 b\|^2. \quad (80)
\]

• Of the two terms on the right we have no control over the second, and we can render the first one zero by solving (the square triangular \( n \times n \) linear system)

\[ R_1 x = Q^T_1 b. \quad (81) \]

• **Definition** (p. 378, textbook): An **inner product** on a vector space \( V \) is a function that, to each pair of vectors \( \mathbf{u} \) and \( \mathbf{v} \) in \( V \), associates a real number \( \langle \mathbf{u}, \mathbf{v} \rangle \) and satisfies the following axioms, for all vectors \( \mathbf{u} \) and \( \mathbf{v} \) in \( V \) and all scalars \( c \):

1. \( \langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle \).
2. \( \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle \)
3. \( \langle c \mathbf{u}, \mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle \)
4. \( \langle \mathbf{u}, \mathbf{u} \rangle \geq 0 \quad \text{and} \quad \langle \mathbf{u}, \mathbf{u} \rangle = 0 \text{ if and only if} \mathbf{u} = 0. \)
• A Vector space with an inner product is called an **inner product space**.

• The Cauchy-Schwarz Inequality says

\[ | <u, v> | \leq \|u\| \|v\| \]  \hspace{1cm} (82)

• The triangle inequality says

\[ \|u + v\| \leq \|u\| + \|v\| . \]  \hspace{1cm} (83)

• One major application of inner product spaces is **weighted least squares**.

• The underlying space is \( \mathbb{R}^n \) and the inner product is

\[ <x, y> = \sum_{i=1}^{n} w_i x_i y_i \]  \hspace{1cm} (84)

where the \( w_i \) are given positive weights.

• The normal equations for the weighted Least Squares Solution of

\[ Ax = b \]  \hspace{1cm} (85)

are

\[ A^T W A x = A^T W b . \]  \hspace{1cm} (86)

• Another major example is **Fourier Series**. The underlying linear space is the set of \( 2\pi \)
periodic functions that are square integrable over an interval of length $2\pi$.

- The underlying inner product is

$$<f, g> = \int_{-\pi}^{\pi} f(t)g(t)dt.$$  \hspace{1cm} (87)

- The Fourier series of a $2\pi$-periodic function $f$ is

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nt) + b_n \sin(nt)$$ \hspace{1cm} (88)

where the Fourier coefficients are given by

$$a_n = \frac{<f, \cos(nt)>}{\pi} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(nt)dt$$

$$b_n = \frac{<f, \sin(nt)>}{\pi} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(nt)dt$$ \hspace{1cm} (89)
4/16/14

Q&A

\[ \| f - F \| = \min \]

2π periodic

\[ \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx + \sum_{k=1}^{\infty} b_k \sin kx \]

\[ = f \]

\[ a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx \]

\[ b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \, dx \]

\[ \langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x) \, dx \]

\[ \| f \| = \sqrt{\langle f, f \rangle} \]
$$A = \begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix}$$

$$|A - 2I| = (2-2)^2 - 6^2$$

$$2 - 2 = 0$$

$$x = 2 \pm 6$$

$$x = 2 + 6$$

$$\begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -6 & 6 \\ 6 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -4 \\ -4 \end{bmatrix} = -4 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = -4 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
\[
\begin{align*}
\left( \begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix} + 4 \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x \\ x \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix} \begin{bmatrix} x \\ x \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} 
\end{align*}
\]

\[\lambda \mathbf{x} = \mathbf{Ax} = 2\mathbf{x}\]

there is some \( i \) in \( \{1, \ldots, n\} \)

\[|a_{ii} - 2| = |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|\]

\[\mathbf{Ax} = \lambda \mathbf{x} \quad \max_{j = 1, \ldots, n} |x_j| = x_i = 1\]

\[
\begin{bmatrix} -3 \\ 2 \end{bmatrix} \rightarrow -\frac{1}{3} \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} = \begin{bmatrix} -1/3 \\ 1 \\ -2/3 \end{bmatrix}
\]

⇒ \text{ith row}

\[
\sum_{j = 1}^{n} a_{ij} x_j = \lambda x_i = \lambda
\]

\[-a_{ii} x_i; \quad \sum_{j \neq i} a_{ij} x_j = \lambda - a_{ii}\]
\[ |\lambda - a_{ii}| = \left| \sum_{j \neq i} a_{ij} x_j \right| \]

\[ \leq \sum_{j \neq i} |a_{ij}| |x_j| \]

\[ = \sum_{j \neq i} |a_{ij}| |x_j| \]

\[ \leq \sum_{j \neq i} |a_{ij}| \]

\[
\begin{bmatrix}
3 \\
2 \\
1
\end{bmatrix}
\]

\[
\begin{bmatrix}
5 \\
6 \\
9
\end{bmatrix}
\]

\[ V_1 = x_1 \]

\[ V_2 = x_2 - \text{proj}_{V_1} x_2 \]
$$V_2 = x_2 - \frac{X_2 \cdot V_1}{V_1 \cdot V_1} V_1$$

$$= \begin{bmatrix} 5 \\ 6 \\ 9 \end{bmatrix} - \frac{36}{14} \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 5 - \frac{54}{7} \\ 6 - \frac{36}{7} \\ 9 - \frac{18}{7} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{5}{7} \\ \frac{6}{7} \\ \frac{9}{7} \end{bmatrix}$$

$$A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = b$$

$$A^T (A \hat{x} - b) = 0$$

$$A^T A \hat{x} = A^T b$$
\[ A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \]

\[ \begin{pmatrix} 14 & 20 \\ 20 & 29 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 9 \\ 13 \end{pmatrix} \]

\[ \begin{pmatrix} a \\ c \\ d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}^{-1} \]