IMPORTANT DEFINITIONS AND THEOREMS

REFERENCE SHEET

This is a (not quite comprehensive) list of definitions and theorems given in Math 1553. Pay particular attention to the ones in red.

Study Tip

For each definition, find an example of something that satisfies the requirements of the definition, and an example of something that does not. For each theorem, find an example of something that satisfies the hypotheses of the theorem, and an example of something that does not satisfy the conclusions (or the hypotheses, of course) of the theorem. This is *great* conceptual practice.

CHAPTER 1

Definition. $\mathbf{R}^n$ = all ordered n-tuples of real numbers $(x_1, x_2, \ldots, x_n)$.

SECTION 1.1.

Definition. A **solution** to a system of linear equations is a list of numbers making *all* of the equations true.

Definition. The **elementary row operations** are the following matrix operations:

- Multiply all entries in a row by a nonzero number (scale).
- Add (a multiple of) each entry of one row to the corresponding entry in another (row replacement).
- Swap two rows.

Definition. Two matrices are called **row equivalent** if one can be obtained from the other by doing some number of elementary row operations.

Definition. A system of equations is called **inconsistent** if it has no solution. It is **consistent** otherwise.

SECTION 1.2.

Definition. A matrix is in **row echelon form** if

(1) All zero rows are at the bottom.
(2) Each leading nonzero entry of a row is to the right of the leading entry of the row above.
(3) Below a leading entry of a row, all entries are zero.

Definition. A **pivot** is the first nonzero entry of a row of a matrix in row echelon form.

Definition. A matrix is in **reduced row echelon form** if it is in row echelon form, and in addition,

(4) The pivot in each nonzero row is equal to 1.
(5) Each pivot is the only nonzero entry in its column.

Theorem. *Every matrix is row equivalent to one and only one matrix in reduced row echelon form.*

Definition. Consider a *consistent* linear system of equations in the variables $x_1, \ldots, x_n$. Let A be the reduced row echelon form of the matrix for this system. We say that x_i is a **free variable** if its corresponding column in A is *not* a pivot column.

Definition. The **parametric form** for the general solution to a system of equations is a system of equations for the non-free variables in terms of the free variables. For instance, if x_2 and x_4 are free,

$$x_1 = 2 - 3x_4 \qquad x_3 = -1 - 4x_4$$

is a parametric form.

Theorem. *Every solution to a consistent linear system is obtained by substituting (unique) values for the free variables in the parametric form.*

Fact. *There are three possibilities for the solution set of a linear system with augmented matrix A:*

(1) *The system is inconsistent: it has zero solutions, and the last column of A is a pivot column.*
(2) *The system has a unique solution: every column of A except the last is a pivot column.*
(3) *The system has infinitely many solutions: the last column isn't a pivot column, and some other column isn't either. These last columns correspond to free variables.*

SECTION 1.3.

Definition. A **point** is an element of $\mathbf{R}^n$, drawn as a point (a dot).
A **vector** is an element of $\mathbf{R}^n$, drawn as an arrow.

Definition. A **scalar** is another name for a real number (to distinguish it from a vector).

Review. Parallelogram law for vector addition.

Definition. A **linear combination** of vectors $v_1, v_2, \ldots, v_n$ is a vector of the form

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

where $c_1, c_2, \ldots, c_n$ are scalars, called the **weights** or **coefficients** of the linear combination.

Definition. A **vector equation** is an equation involving vectors. (It is equivalent to a list of equations involving only scalars. It is also equivalent to an augmented matrix.)

Definition. The **span** of a set of vectors $v_1, v_2, \ldots, v_n$ is the set of all linear combinations of these vectors:

$$\mathrm{Span}\{v_1, \ldots, v_p\} = \left\{ x_1 v_1 + \cdots + x_p v_p \;\middle|\; x_1, \ldots, x_p \text{ in } \mathbf{R} \right\}.$$

Theorem. *The following are equivalent:*

1. *A vector b is in the span of $v_1, v_2, \ldots, v_p$.*
2. *The vector equation*

$$x_1v_1 + x_2v_2 + \cdots + x_pv_p = b$$

has a solution.

3. *The linear system with augmented matrix*

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ v_1 & v_2 & \cdots & v_p & b \\ | & | & & | & | \end{array}\right)$$

is consistent.

SECTION 1.4.

Definition. The **product** of an $m \times n$ matrix A with a vector x in $\mathbf{R}^n$ is the linear combination

$$Ax = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := x_1v_1 + x_2v_2 + \cdots + x_nv_n.$$

The output is a vector in $\mathbf{R}^m$.

Definition. A **matrix equation** is a vector equation involving a product of a matrix with a vector.

Theorem. *$Ax = b$ has a solution if and only if b is in the span of the columns of A.*

Theorem. *Let A be an $m \times n$ (non-augmented) matrix. The following are equivalent*

1. *$Ax = b$ has a solution for all b in $\mathbf{R}^m$.*
2. *The span of the columns of A is all of $\mathbf{R}^m$.*
3. *A has a pivot in each row.*

SECTION 1.5.

Definition. A system of linear equations of the form $Ax = 0$ is called **homogeneous**.

Definition. A system of linear equations of the form $Ax = b$ for $b \neq 0$ is called **inhomogeneous** or **non-homogeneous**.

Definition. The **trivial solution** to a homogeneous equation is the solution $x = 0$: $A0 = 0$.

Theorem. *Let A be a matrix. The following are equivalent:*

1. *$Ax = 0$ has a nontrivial solution.*
2. *There is a free variable.*
3. *A has a column with no pivot.*

Theorem. *The solution set of a homogeneous equation $Ax = 0$ is a span.*

Definition. The **parametric vector form** for the general solution to a system of equations $Ax = b$ is a vector equation expressing all variables in terms of the free variables. For instance, if x_2 and x_4 are free,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ -1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -3 \\ 0 \\ -4 \\ 1 \end{pmatrix}$$

is a parametric vector form. The constant vector $(2, 0, -1, 0)$ is a **specific solution** or **particular solution** to $Ax = b$.

Theorem. *The solution set of a linear system $Ax = b$ is a translate of the solution set of $Ax = 0$ by a specific solution.*

SECTION 1.7.

Definition. A set of vectors $\{v_1, v_2, \dots, v_p\}$ in $\mathbf{R}^n$ is **linearly independent** if the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = 0$$

has only the trivial solution $x_1 = x_2 = \cdots = x_p = 0$.

Definition. A set of vectors $\{v_1, v_2, \dots, v_p\}$ in $\mathbf{R}^n$ is **linearly dependent** if the vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = 0$$

has a nontrivial solution (not all x_i are zero). Such a solution is a **linear dependence relation**.

Theorem. *A set of vectors $\{v_1, v_2, \dots, v_p\}$ is linearly* dependent *if and only if one of the vectors is in the span of the other ones.*

Fact. *Say $v_1, v_2, \dots, v_n$ are in $\mathbf{R}^m$. If $n > m$ then $\{v_1, v_2, \dots, v_n\}$ is linearly* dependent.

Fact. *If one of $v_1, v_2, \dots, v_n$ is zero, then $\{v_1, v_2, \dots, v_n\}$ is linearly* dependent.

Theorem. *Let $v_1, v_2, \dots, v_n$ be vectors in $\mathbf{R}^m$, and let A be the $m \times n$ matrix with columns $v_1, v_2, \dots, v_n$. The following are equivalent:*

1. *The set $\{v_1, v_2, \dots, v_n\}$ is linearly independent.*
2. *No one vector is in the span of the others.*
3. *For every j between 1 and n, v_j is not in* $\mathrm{Span}\{v_1, v_2, \dots, v_{j-1}\}$.
4. *$Ax = 0$ only has the trivial solution.*
5. *A has a pivot in every column.*

Theorem. *Suppose that $\{v_1, v_2, \dots, v_p\}$ is linearly independent. Then every vector w in* $\mathrm{Span}\{v_1, v_2, \dots, v_p\}$ *can be written in* exactly one way *as a linear combination*

$$w = a_1 v_1 + a_2 v_2 + \cdots + a_p v_p.$$

SECTION 1.8.

Definition. A **transformation** (or **function** or **map**) from $\mathbf{R}^n$ to $\mathbf{R}^m$ is a rule T that assigns to each vector x in $\mathbf{R}^n$ a vector $T(x)$ in $\mathbf{R}^m$.

- $\mathbf{R}^n$ is called the **domain** of T.
- $\mathbf{R}^m$ is called the **codomain** of T.
- For x in $\mathbf{R}^n$, the vector $T(x)$ in $\mathbf{R}^m$ is the **image** of x under T.
  Notation: $x \mapsto T(x)$.
- The set of all images $\{T(x) \mid x \text{ in } \mathbf{R}^n\}$ is the **range** of T.

Notation. $T\colon \mathbf{R}^n \longrightarrow \mathbf{R}^m$ means T is a transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$.

Definition. Let A be an $m \times n$ matrix. The **matrix transformation** associated to A is the transformation

$$T\colon \mathbf{R}^n \longrightarrow \mathbf{R}^m \qquad \text{defined by} \qquad T(x) = Ax.$$

- The domain is $\mathbf{R}^n$, where n is the number of columns of A.
- The codomain is $\mathbf{R}^m$, where m is the number of rows of A.
- The range is the span of the columns of A.

Review. Geometric transformations: **projection**, **reflection**, **rotation**, **dilation**, **shear**.

Definition. A **linear transformation** is a transformation T satisfying

$$T(u+v) = T(u) + T(v) \qquad \text{and} \qquad T(cv) = cT(v)$$

for all vectors u, v and all scalars c.

SECTION 1.9.

Definition. The **unit coordinate vectors** in $\mathbf{R}^n$ are

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \ldots, \quad e_{n-1} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Fact. *If A is a matrix, then Ae_i is the ith column of A.*

Definition. Let $T\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. The **standard matrix** for T is

$$\begin{pmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ | & | & & | \end{pmatrix}.$$

Theorem. *If T is a linear transformation, then it is the matrix transformation associated to its standard matrix.*

Definition. A transformation $T\colon \mathbf{R}^n \to \mathbf{R}^m$ is **onto** (or **surjective**) if the range of T is equal to $\mathbf{R}^m$ (its codomain). In other words, each b in $\mathbf{R}^m$ is the image of *at least one* x in $\mathbf{R}^n$.

Theorem. *Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation with matrix A. Then the following are equivalent:*

- *T is onto*
- *$T(x) = b$ has a solution for every b in $\mathbf{R}^m$*
- *$Ax = b$ is consistent for every b in $\mathbf{R}^m$*
- *The columns of A span $\mathbf{R}^m$*
- *A has a pivot in every row.*

Definition. A transformation $T : \mathbf{R}^n \to \mathbf{R}^m$ is **one-to-one** (or **into**, or **injective**) if different vectors in $\mathbf{R}^n$ map to different vectors in $\mathbf{R}^m$. In other words, each b in $\mathbf{R}^m$ is the image of *at most one* x in $\mathbf{R}^n$.

Theorem. *Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation with matrix A. Then the following are equivalent:*

- *T is one-to-one*
- *$T(x) = b$ has one or zero solutions for every b in $\mathbf{R}^m$*
- *$Ax = b$ has a unique solution or is inconsistent for every b in $\mathbf{R}^m$*
- *$Ax = 0$ has a unique solution*
- *The columns of A are linearly independent*
- *A has a pivot in every column.*

Chapter 2

Section 2.1.

Definition. The ij**th entry** of a matrix A is the entry in the ith row and jth column. Notation: a_{ij}.

Definition. The entries $a_{11}, a_{22}, a_{33}, \ldots$ are the **diagonal entries**; they form the **main diagonal** of the matrix.

Definition. A **diagonal matrix** is a *square* matrix whose only nonzero entries are on the main diagonal.

Definition. The $n \times n$ **identity matrix** I_n is the diagonal matrix with all diagonal entries equal to 1. It has the property that $I_n A = A$ for any $n \times m$ matrix A.

Definition. The **zero matrix** (of size $m \times n$) is the $m \times n$ matrix 0 with all zero entries.

Definition. The **transpose** of an $m \times n$ matrix A is the $n \times m$ matrix A^T whose rows are the columns of A. In other words, the ij entry of A^T is a_{ji}.

Definition. The **product** of an $m \times n$ matrix A with an $n \times p$ matrix B is the $m \times p$ matrix

$$AB = \begin{pmatrix} | & | & & | \\ Av_1 & Av_2 & \cdots & Av_p \\ | & | & & | \end{pmatrix},$$

where $v_1, v_2, \ldots, v_p$ are the columns of B.

Fact. *Suppose A has is an $m \times n$ matrix, and that the other matrices below have the right size to make multiplication work. Then:*

$$\begin{aligned} A(BC) &= (AB)C & \qquad A(B+C) &= AB+AC \\ (B+C)A &= BA+CA & c(AB) &= (cA)B \\ c(AB) &= A(cB) & I_nA &= A \\ AI_m &= A \end{aligned}$$

Fact. *If A, B, and C are matrices, then:*

(1) *AB is usually not equal to BA.*
(2) *$AB = AC$ does not imply $B = C$.*
(3) *$AB = 0$ does not imply $A = 0$ or $B = 0$.*

Definition. Let $T: \mathbf{R}^n \to \mathbf{R}^m$ and $U: \mathbf{R}^p \to \mathbf{R}^n$ be transformations. The **composition** is the transformation

$$T \circ U : \mathbf{R}^p \to \mathbf{R}^m \quad \text{defined by} \quad T \circ U(x) = T(U(x)).$$

Theorem. *Let $T: \mathbf{R}^n \to \mathbf{R}^m$ and $U: \mathbf{R}^m \to \mathbf{R}^p$ be linear transformations with matrices A and B, respectively. Then the matrix for $T \circ U$ is AB.*

SECTION 2.2.

Definition. A square matrix A is **invertible** (or **nonsingular**) if there is a matrix B of the same size, such that

$$AB = I_n \quad \text{and} \quad BA = I_n.$$

In this case we call B the **inverse** of A, and we write $A^{-1} = B$.

Theorem. *If A is invertible, then $Ax = b$ has exactly one solution for every b, namely:*

$$x = A^{-1}b.$$

Fact. *Suppose that A and B are invertible $n \times n$ matrices.*

(1) *A^{-1} is invertible and its inverse is $(A^{-1})^{-1} = A$.*
(2) *AB is invertible and its inverse is $(AB)^{-1} = B^{-1}A^{-1}$.*
(3) *A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.*

Theorem. *Let A be an $n \times n$ matrix. Here's how to compute A^{-1}.*

(1) *Row reduce the augmented matrix $(\, A \mid I_n \,)$.*
(2) *If the result has the form $(\, I_n \mid B \,)$, then A is invertible and $B = A^{-1}$.*
(3) *Otherwise, A is not invertible.*

Theorem. *An $n \times n$ matrix A is invertible if and only if it is row equivalent to I_n. In this case, the sequence of row operations taking A to I_n also takes I_n to A^{-1}.*

Definition. The **determinant** of a 2×2 matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is

$$\det(A) = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Fact. *If A is a 2×2 matrix, then A is invertible if and only if* $\det(A)\neq 0$. *In this case,*

$$A^{-1}=\frac{1}{\det(A)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Definition. A **elementary matrix** is a square matrix E which differs from the identity matrix by exactly one row operation.

Fact. *If E is the elementary matrix for a row operation, and A is a matrix, then EA differs from A by the same row operation.*

SECTION 2.3.

Definition. A transformation $T\colon \mathbf{R}^n \to \mathbf{R}^n$ is **invertible** if there exists another transformation $U\colon \mathbf{R}^n \to \mathbf{R}^n$ such that

$$T\circ U(x)=x \quad\text{and}\quad U\circ T(x)=x$$

for all x in $\mathbf{R}^n$. In this case we say U is the **inverse** of T, and we write $U=T^{-1}$.

Fact. *A transformation T is invertible if and only if it is both one-to-one and onto.*

Theorem. *If T is an invertible linear transformation with matrix A, then T^{-1} is an invertible linear transformation with matrix A^{-1}.*

I'll keep all of the conditions of the IMT right here, even though we don't encounter some until later:

The Invertible Matrix Theorem. *Let A be a square $n\times n$ matrix, and let $T\colon \mathbf{R}^n \to \mathbf{R}^n$ be the linear transformation $T(x)=Ax$. The following statements are equivalent.*

(1) *A is invertible.*
(2) *T is invertible.*
(3) *A is row equivalent to I_n.*
(4) *A has n pivots.*
(5) *$Ax=0$ has only the trivial solution.*
(6) *The columns of A are linearly independent.*
(7) *T is one-to-one.*
(8) *$Ax=b$ is consistent for all b in $\mathbf{R}^n$.*
(9) *The columns of A span $\mathbf{R}^n$.*
(10) *T is onto.*
(11) *A has a left inverse (there exists B such that $BA=I_n$).*
(12) *A has a right inverse (there exists B such that $AB=I_n$).*
(13) *A^T is invertible.*
(14) *The columns of A form a basis for $\mathbf{R}^n$.*
(15) $\operatorname{Col}A=\mathbf{R}^n$.
(16) $\dim\operatorname{Col}A=n$.
(17) $\operatorname{rank}A=n$.
(18) $\operatorname{Nul}A=\{0\}$.
(19) $\dim\operatorname{Nul}A=0$.
(20) $\det(A)\neq 0$.

(21) *The number* 0 *is* not *an eigenvalue of* A*.*

SECTION 2.8.

Definition. A **subspace** of $\mathbf{R}^n$ is a subset V of $\mathbf{R}^n$ satisfying:

(1) The zero vector is in V.
(2) If u and v are in V, then $u+v$ is also in V.
(3) If u is in V and c is in $\mathbf{R}$, then cu is in V.

Definition. If $V = \mathrm{Span}\{v_1, v_2, \ldots, v_n\}$, we say that V is the subspace **generated by** or **spanned by** the vectors $v_1, v_2, \ldots, v_n$.

Theorem. *A subspace is a span, and a span is a subspace.*

Definition. The **column space** of a matrix A is the subspace spanned by the columns of A. It is written $\mathrm{Col}\,A$.

Definition. The **null space** of A is the set of all solutions of the homogeneous equation $Ax = 0$:

$$\mathrm{Nul}\,A = \big\{x \mid Ax = 0\big\}.$$

Example. The following are the most important examples of subspaces in this class (some won't appear until later):

- Any $\mathrm{Span}\{v_1, v_2, \ldots, v_m\}$.
- The column space of a matrix: $\mathrm{Col}\,A = \mathrm{Span}\{\text{columns of } A\}$.
- The range of a linear transformation (same as above).
- The null space of a matrix: $\mathrm{Nul}\,A = \big\{x \mid Ax = 0\big\}$.
- The row space of a matrix: $\mathrm{Row}\,A = \mathrm{Span}\{\text{rows of } A\}$.
- The λ-eigenspace of a matrix, where λ is an eigenvalue.
- The orthogonal complement $W^\perp$ of a subspace W.
- The zero subspace $\{0\}$.
- All of $\mathbf{R}^n$.

Definition. Let V be a subspace of $\mathbf{R}^n$. A **basis** of V is a set of vectors $\{v_1, v_2, \ldots, v_m\}$ in V such that:

(1) $V = \mathrm{Span}\{v_1, v_2, \ldots, v_m\}$, and
(2) $\{v_1, v_2, \ldots, v_m\}$ is linearly independent.

The number of vectors in a basis is the **dimension** of V, and is written $\dim V$.

Theorem. *Every basis for a gives subspace has the same number of vectors in it.*

Fact. *The vectors in the parametric vector form of the general solution to* $Ax = 0$ *always form a basis for* $\mathrm{Nul}\,A$*.*

Fact. *The pivot columns of* A *always form a basis for* $\mathrm{Col}\,A$*.*

SECTION 2.9.

Definition. Let $\mathcal{B} = \{v_1, v_2, \ldots, v_m\}$ be a basis of a subspace V. Any vector x in V can be written uniquely as a linear combination $x = c_1v_1 + c_2v_2 + \cdots + c_mv_m$. The coefficients $c_1, c_2, \ldots, c_m$ are the **coordinates of x with respect to $\mathcal{B}$**, and the vector with entries $c_1, c_2, \ldots, c_m$ is the **$\mathcal{B}$-coordinate vector of x**, denoted $[x]_{\mathcal{B}}$. In summary,

$$[x]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} \qquad \text{means} \qquad x = c_1v_1 + c_2v_2 + \cdots + c_mv_m.$$

Definition. The **rank** of a matrix A, written $\operatorname{rank} A$, is the dimension of the column space $\operatorname{Col} A$.

Rank Theorem. *If A is an $m \times n$ matrix, then*

$$\operatorname{rank} A + \dim \operatorname{Nul} A = n = \textit{the number of columns of } A.$$

Basis Theorem. *Let V be a subspace of dimension m. Then:*

- *Any m linearly independent vectors in V form a basis for V.*
- *Any m vectors that span V form a basis for V.*

CHAPTER 3

Definition. The **determinant** is a function

$$\det\colon \{\text{square matrices}\} \longrightarrow \mathbf{R}$$

with the following **defining properties**:

(1) If we do a row replacement on a matrix, the determinant does not change.
(2) If we scale a row of a matrix by k, the determinant scales by k.
(3) If we swap two rows of a matrix, the determinant scales by -1.
(4) $\det(I_n) = 1$

Magical Properties of the Determinant.

(1) *There is one and only one function* $\det\colon \{\textit{square matrices}\} \to \mathbf{R}$ *satisfying the defining properties (1)–(4).*
(2) *A is invertible if and only if* $\det(A) \neq 0$.
(3) *If we row reduce A to row echelon form B using r swaps, then*

$$\det(A) = (-1)^r \frac{\text{(product of the diagonal entries of } B\text{)}}{\text{(product of the scaling factors)}}.$$

(4) $\det(AB) = \det(A)\det(B)$ *and* $\det(A^{-1}) = \det(A)^{-1}$
(5) $\det(A) = \det(A^T)$
(6) $|\det(A)|$ *is the volume of the parallelepiped defined by the columns of A.*
(7) *If A is an $n \times n$ matrix with transformation $T(x) = Ax$, and S is a subset of $\mathbf{R}^n$, then the volume of $T(S)$ is* $|\det(A)|$ *times the volume of S. (Even for curvy shapes S.)*

Definition. The ij **minor** of an $n \times n$ matrix A is the $(n-1) \times (n-1)$ matrix A_{ij} you get by deleting the ith row and the jth column from A.

Definition. The ij **cofactor** of A is $C_{ij} = (-1)^{i+j} \det A_{ij}$.

Theorem. *The determinant of an $n \times n$ matrix A can be calculated using **cofactor expansion** along any row or column:*

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij} \text{ for any fixed } i$$

$$\det A = \sum_{i=1}^{n} a_{ij} C_{ij} \text{ for any fixed } j$$

Theorem. *There are special formulas for determinants of 2×2 and 3×3 matrices:*

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{array}{r} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \end{array}$$

Theorem. *The determinant of an upper-triangular or lower-triangular matrix is the product of the diagonal entries.*

Theorem. *If A is an invertible $n \times n$ matrix, then*

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} C_{11} & C_{21} & C_{31} & \cdots & C_{n1} \\ C_{12} & C_{22} & C_{32} & \cdots & C_{n2} \\ C_{13} & C_{23} & C_{33} & \cdots & C_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & C_{3n} & \cdots & C_{nn} \end{pmatrix}$$

CHAPTER 5

SECTION 5.1.

Definition. Let A be an $n \times n$ matrix.

(1) An **eigenvector** of A is a nonzero vector v in $\mathbf{R}^n$ such that $Av = \lambda v$, for some λ in $\mathbf{R}$. In other words, Av is a multiple of v.
(2) An **eigenvalue** of A is a number λ in $\mathbf{R}$ such that the equation $Av = \lambda v$ has a nontrivial solution.

If $Av = \lambda v$ for $v \neq 0$, we say λ is the **eigenvalue for** v, and v is an **eigenvector for** λ.

Fact. *The eigenvalues of a triangular matrix are the diagonal entries.*

Fact. *A matrix is invertible if and only if zero is not an eigenvalue.*

Fact. *Eigenvectors with distinct eigenvalues are linearly independent.*

Definition. Let A be an $n\times n$ matrix and let λ be an eigenvalue of A. The **λ-eigenspace** of A is the set of all eigenvectors of A with eigenvalue λ, plus the zero vector:

$$\begin{aligned}\lambda\text{-eigenspace} &= \{v \text{ in } \mathbf{R}^n \mid Av = \lambda v\} \\ &= \{v \text{ in } \mathbf{R}^n \mid (A-\lambda I)v = 0\} \\ &= \mathrm{Nul}(A-\lambda I).\end{aligned}$$

SECTION 5.2.

Definition. Let A be an $n\times n$ matrix. The **characteristic polynomial** of A is

$$f(\lambda) = \det(A-\lambda I).$$

The **characteristic equation** of A is the equation

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

Fact. *If A is an $n\times n$ matrix, then the characteristic polynomial of A has degree n.*

Fact. *The roots of the characteristic polynomial (i.e., the solutions of the characteristic equation) are the eigenvalues of A.*

Definition. The **algebraic multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic polynomial.

Definition. Two $n\times n$ matrices A and B are **similar** if there is an invertible $n\times n$ matrix C such that $A = CBC^{-1}$.

Fact. *Similar matrices have the same characteristic polynomial, hence the same eigenvalues (but different eigenvectors in general).*

SECTION 5.3.

Definition. An $n\times n$ matrix A is **diagonalizable** if it is similar to a diagonal matrix:

$$A = PDP^{-1} \qquad \text{for } D \text{ diagonal}.$$

Fact. *If $A = PDP^{-1}$ for $D = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn}\end{pmatrix}$, then*

$$A^m = PD^mP^{-1} = P\begin{pmatrix} d_{11}^m & 0 & \cdots & 0 \\ 0 & d_{22}^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn}^m\end{pmatrix}P^{-1}.$$

The Diagonalization Theorem. *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. In this case, $A = PDP^{-1}$ for*

$$P = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \qquad D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where $v_1, v_2, \ldots, v_n$ are linearly independent eigenvectors, and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the corresponding eigenvalues (in the same order).

Corollary. *An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.*

Procedure. *How to diagonalize a matrix A:*

(1) Find the eigenvalues of A using the characteristic polynomial.
(2) For each eigenvalue λ of A, compute a basis $\mathcal{B}_\lambda$ for the λ-eigenspace.
(3) If there are fewer than n total vectors in the union of all of the eigenspaces $\mathcal{B}_\lambda$, then the matrix is not diagonalizable.
(4) Otherwise, the n vectors $v_1, v_2, \ldots, v_n$ in your eigenspace bases are linearly independent, and $A = PDP^{-1}$ for

$$P = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where λ_i is the eigenvalue for v_i.

Definition. Let λ be an eigenvalue of a square matrix A. The **geometric multiplicity** of λ is the dimension of the λ-eigenspace.

Theorem. *Let λ be an eigenvalue of a square matrix A. Then*

$$1 \leq (\text{the geometric multiplicity of } \lambda) \leq (\text{the algebraic multiplicity of } \lambda).$$

Corollary. *Let λ be an eigenvalue of a square matrix A. If the algebraic multiplicity of λ is* 1*, then the geometric multiplicity is also* 1.

The Diagonalization Theorem (Alternate Form). *Let A be an $n \times n$ matrix. The following are equivalent:*

(1) *A is diagonalizable.*
(2) *The sum of the geometric multiplicities of the eigenvalues of A equals n.*
(3) *The sum of the algebraic multiplicities of the eigenvalues of A equals n, and* the geometric multiplicity equals the algebraic multiplicity *of each eigenvalue.*

SECTION 5.5.

Review. Arithmetic in the complex numbers.

The Fundamental Theorem of Algebra. *Every polynomial of degree n has exactly n complex roots, counted with multiplicity.*

Fact. *Complex roots of real polynomials come in* conjugate pairs.

Fact. *If λ is an eigenvalue of a real matrix with eigenvector v, then $\overline{\lambda}$ is also an eigenvalue, with eigenvector $\overline{v}$.*

Theorem. *Let A be a 2×2 matrix with complex (non-real) eigenvalue λ, and let v be an eigenvector. Then*

$$A = PCP^{-1}$$

where

$$P = \begin{pmatrix} | & | \\ \operatorname{Re} v & \operatorname{Im} v \\ | & | \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \operatorname{Re}\lambda & \operatorname{Im}\lambda \\ -\operatorname{Im}\lambda & \operatorname{Re}\lambda \end{pmatrix}.$$

The matrix C is a composition of rotation by $-\arg(\lambda)$ and scaling by $|\lambda|$.

Theorem. *Let A be a real $n\times n$ matrix. Suppose that for each (real or complex) eigenvalue, the dimension of the eigenspace equals the algebraic multiplicity. Then $A = PCP^{-1}$, where P and C are as follows:*

(1) *C is **block diagonal**, where the blocks are 1×1 blocks containing the real eigenvalues (with their multiplicities), or 2×2 blocks containing the matrices $\begin{pmatrix} \operatorname{Re}\lambda & \operatorname{Im}\lambda \\ -\operatorname{Im}\lambda & \operatorname{Re}\lambda \end{pmatrix}$ for each complex eigenvalue λ (with multiplicity).*

(2) *The columns of P form bases for the eigenspaces for the real eigenvectors, or come in pairs $(\operatorname{Re} v \ \ \operatorname{Im} v)$ for the complex eigenvectors.*

CHAPTER 6

SECTION 6.1.

Definition.
The **dot product** of two vectors x, y in $\mathbf{R}^n$ is

$$x \cdot y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \overset{\text{def}}{=} x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

Thinking of x, y as column vectors, this is the same as the number $x^T y$.

Definition. The **length** or **norm** of a vector x in $\mathbf{R}^n$ is

$$\|x\| = \sqrt{x \cdot x}.$$

Fact. *If x is a vector and c is a scalar, then $\|cx\| = |c| \cdot \|x\|$.*

Definition. The **distance** between two points x, y in $\mathbf{R}^n$ is

$$\mathrm{dist}(x, y) = \|y - x\|.$$

Definition. A **unit vector** is a vector v with length $\|v\| = 1$.

Definition. Let x be a nonzero vector in $\mathbf{R}^n$. The **unit vector in the direction of** x is the vector $x/\|x\|$.

Definition. Two vectors x, y are **orthogonal** or **perpendicular** if $x \cdot y = 0$.
Notation: $x \perp y$.

Fact. $x \perp y \iff \|x - y\|^2 = \|x\|^2 + \|y\|^2$

Definition. Let W be a subspace of $\mathbf{R}^n$. Its **orthogonal complement** is

$$W^\perp = \big\{ v \text{ in } \mathbf{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W \big\}.$$

Fact. *Let W be a subspace of $\mathbf{R}^n$.*

(1) *$W^\perp$ is also a subspace of $\mathbf{R}^n$*
(2) $(W^\perp)^\perp = W$
(3) $\dim W + \dim W^\perp = n$
(4) *If $W = \mathrm{Span}\{v_1, v_2, \ldots, v_m\}$, then*

$$\begin{aligned} W^\perp &= \textit{all vectors orthogonal to each } v_1, v_2, \ldots, v_m \\ &= \big\{ x \textit{ in } \mathbf{R}^n \mid x \cdot v_i = 0 \textit{ for all } i = 1, 2, \ldots, m \big\} \\ &= \mathrm{Nul}\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}. \end{aligned}$$

Definition. The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted $\mathrm{Row}A$. Equivalently, it is the column span of A^T:

$$\mathrm{Row}A = \mathrm{Col}A^T.$$

It is a subspace of $\mathbf{R}^n$.

Fact. $\mathrm{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \mathrm{Nul}\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}.$

Fact. *Let A be a matrix.*

(1) $(\mathrm{Row}A)^\perp = \mathrm{Nul}A$ *and* $(\mathrm{Nul}A)^\perp = \mathrm{Row}A$.
(2) $(\mathrm{Col}A)^\perp = \mathrm{Nul}A^T$ *and* $(\mathrm{Nul}A^T)^\perp = \mathrm{Col}A$.

SECTIONS 6.2 AND 6.3.

Theorem. *Let W be a subspace of $\mathbf{R}^n$. Every vector x can be decompsed uniquely as*

$$x = x_W + x_{W^\perp}$$

where x_W is the closest vector to x in W, and $x_{W^\perp}$ is in $W^\perp$.

The equation $x = x_W + x_{W^\perp}$ is called the **orthogonal decomposition** of x with respect to W.

Definition. A set of *nonzero* vectors is **orthogonal** if each pair of vectors is orthogonal. It is **orthonormal** if, in addition, each vector is a unit vector.

Lemma. *A set of orthogonal vectors is linearly independent. Hence it is a basis for its span.*

Definition. Let W be a subspace of $\mathbf{R}^n$, and let $\{u_1, u_2, \ldots, u_m\}$ be an *orthogonal* basis for W. The **orthogonal projection** of a vector x onto W is

$$\mathrm{proj}_W(x) \stackrel{\text{def}}{=} \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{x \cdot u_n}{u_n \cdot u_n} u_n.$$

For instance, if $L = \mathrm{Span}\{u\}$ is a line, then

$$\mathrm{proj}_L(x) = \frac{x \cdot u}{u \cdot u} u.$$

Theorem. *Let W be a subspace of $\mathbf{R}^n$, and let x be a vector in $\mathbf{R}^n$. Then*

$$x_W = \mathrm{proj}_W(x) \quad \textit{and} \quad x_{W^\perp} = x - \mathrm{proj}_W(x).$$

In particular, $\mathrm{proj}_W(x)$ is the closest point in W to x:

Best Approximation Theorem. *Let W be a subspace of $\mathbf{R}^n$, and let x be a vector in $\mathbf{R}^n$. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that*

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \textit{for all} \quad y' \textit{ in } W.$$

Definition. We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W : \mathbf{R}^n \longrightarrow \mathbf{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

Theorem. *Let W be a subspace of $\mathbf{R}^n$.*

(1) proj_W *is a* linear *transformation.*
(2) *For every x in W, we have $\mathrm{proj}_W(x) = x$.*
(3) *For every x in $W^\perp$, we have $\mathrm{proj}_W(x) = 0$.*
(4) *The range of proj_W is W and the null space of proj_W is $W^\perp$.*

Corollary. *Let $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ be an orthogonal set, and let x be a vector in $W = \mathrm{Span}\,\mathcal{B}$. Then*

$$x = \mathrm{proj}_W(x) = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m} u_m.$$

In other words, the $\mathcal{B}$-coordinates of x are $\left(\frac{x \cdot u_1}{u_1 \cdot u_1}, \frac{x \cdot u_2}{u_2 \cdot u_2}, \ldots, \frac{x \cdot u_m}{u_1 \cdot u_m} \right)$.

Fact. *Let W be an m-dimensional subspace of $\mathbf{R}^n$, let $\operatorname{proj}_W : \mathbf{R}^n \to \mathbf{R}^n$ be the projection, and let A be the matrix for proj_L.*

(1) *A is diagonalizable with eigenvalues 0 and 1; it is similar to the diagonal matrix with m ones and $n-m$ zeros on the diagonal.*
(2) *$A^2 = A$.*

SECTION 6.4.

The Gram–Schmidt Process. *Let $\{v_1, v_2, \ldots, v_m\}$ be a basis for a subspace W of $\mathbf{R}^n$. Define:*

(1) $u_1 = v_1$
(2) $u_2 = v_2 - \operatorname{proj}_{\operatorname{Span}\{u_1\}}(v_2) \qquad = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1$
(3) $u_3 = v_3 - \operatorname{proj}_{\operatorname{Span}\{u_1,u_2\}}(v_3) \qquad = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2$

$\vdots$

m. $u_m = v_m - \operatorname{proj}_{\operatorname{Span}\{u_1,u_2,\ldots,u_{m-1}\}}(v_m) \quad = v_m - \displaystyle\sum_{i=1}^{m-1} \frac{v_m \cdot u_i}{u_i \cdot u_i} u_i$

Then $\{u_1, u_2, \ldots, u_m\}$ is an orthogonal *basis for the same subspace W.*

SECTION 6.5.

Definition. A **least squares solution** to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

$$\|b - A\widehat{x}\| \leq \|b - Ax\|$$

for all x in $\mathbf{R}^n$.

Theorem. *The least squares solutions to $Ax = b$ are the solutions to*

$$(A^T A)\widehat{x} = A^T b.$$

Theorem. *If A has orthogonal columns $v_1, v_2, \ldots, v_n$, then the least squares solution to $Ax = b$ is*

$$\widehat{x} = \left(\frac{b \cdot v_1}{v_1 \cdot v_1}, \frac{b \cdot v_2}{v_2 \cdot v_2}, \cdots, \frac{b \cdot v_n}{v_n \cdot v_n} \right).$$

Theorem. *Let A be an $m \times n$ matrix. The following are equivalent:*

(1) *$Ax = b$ has a* unique *least squares solution for all b in $\mathbf{R}^n$.*
(2) *The columns of A are linearly independent.*
(3) *$A^T A$ is invertible.*

In this case, the least squares solution is $(A^T A)^{-1}(A^T b)$.

Review. Examples of best fit problems using least squares.