Lecture Notes on Linear Algebra

Érik Amorim
## Contents

**Introduction** 5

**Chapter 1. Matrices, vectors, systems of linear equations** 7
  - 1.1. Matrices and vectors 7
  - 1.2. Linear combinations, matrix-vector products, some special matrices 10
  - 1.3. Systems of linear equations 13
  - 1.4. Gaussian Elimination 17
  - 1.5. Applications of linear systems 19
  - 1.6. Span 19
  - 1.7. Linear dependence and independence 21

**Chapter 2. Matrices and Linear Transformations** 25
  - 2.1. Matrix multiplication 25
  - 2.2. Applications of matrix products 27
  - 2.3. Invertibility and elementary matrices 28
  - 2.4. Matrix inversion 32
  - 2.5. Block multiplication 33
  - 2.6. LU Decomposition 33
  - 2.7. Linear transformations and matrices 35
  - 2.8. Composition and invertibility 38

**Chapter 3. Determinants** 41
  - 3.1. Cofactor expansion 41
  - 3.2. Properties of determinants 44

**Chapter 4. Linear subspaces** 47
  - 4.1. Linear subspaces 47
  - 4.2. Basis and dimension 49
  - 4.3. Dimension of subspaces associated to a matrix 52
  - 4.4. Coordinate systems 54
  - 4.5. Matrix representations of linear transformations in other bases 57

**Chapter 5. Eigenstuff** 61
  - 5.1. Eigenvalues and eigenvectors 61
  - 5.2. Characteristic polynomial 65
  - 5.3. Diagonalization 68
  - 5.4. Diagonalization of operators 71
  - 5.5. Applications of diagonalization 71

**Chapter 6. Orthogonality** 77
  - 6.1. Geometry of vectors 77
  - 6.2. Orthogonal vectors 81
  - 6.3. Orthogonal projections 83
  - 6.4. Least Squares Algorithm 86
  - 6.5. Orthogonal matrices 89
  - 6.6. Symmetric matrices 91
Introduction

What is Linear Algebra? In high school you likely studied systems of linear equations and solved them using different methods. But sometimes you find a unique solution, sometimes infinitely many (there are free variables) and sometimes no solution at all (there are inconsistencies in the equations). In this class we will build up the theoretical framework underlying this phenomenon, which is based on **vectors** and **matrices**, and study simple algorithms that can in some sense quantify the degree to which a system is inconsistent or has infinitely many solutions. But vectors are naturally geometric objects, and matrices can act on them by transforming them into other vectors, so it turns out that these ideas have wide applicability for both theoretical and practical purposes in different areas of science involving either geometry or algebra, and the unifying concept behind it all is that of a **linear transformation**.

In this class, we will study both the theoretical mathematics and the algorithms provided by the theory of linear transformations, using the language of matrices and vectors, with the emphasis on being able to apply the algorithms, but not neglecting the understanding of the theory.

These notes are based on the book *Elementary Linear Algebra: A Matrix Approach*, by Spence, Insel & Friedberg, 2nd edition (Prentice Hall 2017, ISBN 9780134689470). I changed the names of chapters and sections, but the organization of the material in chapters and sections is mostly the same as in the book, apart from small changes in the order of presentation in some sections.

The role of proofs in this class. These notes are not presented as a self-contained, logically consistent treatment of linear algebra, like most math books. Sometimes proofs are provided, sometimes not. Sometimes concepts are introduced in an order that would not make sense if you wanted to prove that they are well-defined. This is done whenever I feel like what’s more important about them is to be able to compute with them, but not necessarily to know why they are computed this way.

Students are expected to be able to reproduce some of the simple proofs given here and come up with their own for similar problems. There are suggested problems from the book indicated. But they are not expected to be able to understand the nuances of proof writing and formal mathematical reasoning, which at Rutgers are introduced in Math 300. Only those proof-based problems where it’s clear how to proceed will be tested, not those requiring original, creative ideas.

Content organization. I organized the content in these notes in boxes followed by examples where pertinent. Most boxes have a label that indicate their role in the text:

- **DEFINITION:** New concepts are introduced. Students should be able to reproduce these definitions in writing as precise mathematical statements.
- **PROBLEM and SOLUTION:** Essentially an example, but one that can be formulated as exactly the type of problems that might be encountered for that section.
- **THEOREM:** A result that’s not trivial to prove. It may be followed by a *proof*, either immediately or after a few examples, but some proofs have been omitted. Most results that could also be called theorems are presented in unlabelled boxes, because they should not be thought of as hard to understand.
- **PROPERTIES:** Like a theorem, but listing several properties that the object in question satisfies or doesn’t satisfy.
- **ALGORITHM:** A procedure that can be followed as stated to calculate whatever object it calculates. It may be followed by a *justification*.
- **Justification:** An informally presented proof for why some statement is true or some algorithm works. For example, it may not cover all possible cases that need to be analyzed or it may only give a reason that suggests the result could be true, like a proof by example.
- **Proof:** A mathematically sound proof of a previous statement or theorem, even if presented with informal or condescending language.
- **REMARK:** Some extra words about cases or features not included in the text. They are mentioned for completeness, but will not be tested on exams or homeworks.
CHAPTER 1

Matrices, vectors, systems of linear equations

Matrices are $m \times n$ tables of numbers. Vectors are columns of $m$ numbers that can naturally be understood as points in $m$-dimensional space, just like a point in ordinary 3D space can be thought of as a vector containing its $x$, $y$ and $z$ coordinates. Matrices can multiply vectors to produce new vectors. Linear systems can also be cast into the language of matrices and vectors, and the simple method of elimination for solving a system will be made precise and efficient under the name of Gaussian Elimination, which is an algorithm that you absolutely must master because it will be used again and again. It brings a matrix into a simple form called its RREF form, which allows one to quickly solve the linear system at hand, but also to understand and quantify its consistency and unique solvability, which will go by the names of rank and nullity.

Then we come to a somewhat unrelated but very important idea: a collection of a few vectors is said to generate (or span) a subspace, which could even be the entire $m$-dimensional space, by linear combinations that can be formed from them. We will learn how to determine whether you really needed each of those vectors to span it or if there was redundancy (linear dependence) among them. Identifying a set of vectors that span some given subspace allows us to understand properties of that subspace.

1.1. Matrices and vectors

Key concepts: basic matrix arithmetic, matrix transpose, vectors

**DEFINITION (matrix)**

- **Matrix:** A table of numbers, denoted with an uppercase letter ($A, B, C, ...$).
- **Dimensions:** An $m \times n$ matrix (notation: $A_{m \times n}$) is a matrix with $m$ rows and $n$ columns.
- **Entries or elements:** The $(i,j)$-th entry of $A$ (notation: $a_{ij}$) is the number located at row $i$ and column $j$ of the matrix.

We enclose the entries of a matrix by parentheses $( )$ or square brackets $[ ]$.

- $A = \begin{bmatrix} 1 & 2 & 3 \\ -\frac{1}{2} & \pi & 0 \end{bmatrix}$ is a $2 \times 3$ matrix with entries $\begin{cases} a_{11} = 1 \\ a_{12} = 2 \\ a_{13} = 3 \\ a_{21} = -\frac{1}{2} \\ a_{22} = \pi \\ a_{23} = 0 \end{cases}$.
- $B = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$ is a $3 \times 1$ matrix whose entries all satisfy $b_{ij} = 4$ (for $1 \leq i \leq 3$ and $j = 1$).
- $C = [ -6 ]$ is a $1 \times 1$ matrix whose only entry is $c_{11} = -6$.
- The matrix $D$ described by the rule $d_{ij} = i + j$ for $1 \leq i \leq 2$ and $1 \leq j \leq 5$ is

$$D = \begin{bmatrix} 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 6 & 7 \end{bmatrix}$$

That is, its entries are given by the sum of their row and column numbers.

**DEFINITION (scalar multiplication)**

Matrices can be multiplied by a scalar (a number) elementwise.

We generally use letters $r, s, t, ...$ for scalars.

The scalar is written on the left of the matrix: $rA$ instead of $Ar$.

- For the matrix $A = \begin{bmatrix} 1 & 0 \\ 2 & -5 \end{bmatrix}$, we have

$$3A = \begin{bmatrix} 3 & 0 \\ 6 & -15 \end{bmatrix} - A = \begin{bmatrix} -1 & 0 \\ -2 & 5 \end{bmatrix} 0A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$
DEFINITION (matrix addition)

Matrices of the same dimensions can be summed (or subtracted) elementwise.

Zero matrix: The zero matrix of a given dimension $m \times n$, denoted $O$ or $O_{m \times n}$, is the matrix with zeroes for all entries.

- For $A = \begin{bmatrix} 1 & 1 \\ 0 & 3 \\ -2 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 1 \\ -1 & 0 \\ -1 & 3 \end{bmatrix}$, we have
  \[ A + B = \begin{bmatrix} 0 & 2 \\ -1 & 3 \\ -3 & 8 \end{bmatrix} \quad 2A - B = \begin{bmatrix} 3 & 1 \\ -1 & 6 \\ -3 & 7 \end{bmatrix} \]

- For $C = \begin{bmatrix} 3 & 2 & 1 \end{bmatrix}$, we have $C + O_{1 \times 3} = \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} = C$.
- The matrix sum below (asterisks represent any numbers)
  \[ \begin{bmatrix} * & * \\ * & * \end{bmatrix} + \begin{bmatrix} * & * \\ * & * \end{bmatrix} \]
  is undefined because the dimensions are not the same.

PROPERTIES (matrix addition and scalar multiplication)

The following properties are intuitive. A formal proof would likely confuse more than it would provide understanding. See Theorem 1.1 on page 6.

Let $A, B$ be matrices of the same dimensions, let $r, s$ be scalars. Then:

- Commutative law: $A + B = B + A$
- Associative law: $(A + B) + C = A + (B + C)$
- Distributive laws: $(r + s)A = rA + sA$ and $r(A + B) = rA + rB$
- Identity element: $O + A = A$

DEFINITION (transpose)

The transpose of an $m \times n$ matrix $A$ is the $n \times m$ matrix $A^T$ whose columns are the rows of $A$ and vice-versa. The $(i, j)$-th element of $A^T$ (we can denote it by $a^T_{ij}$) is equal to $a_{ji}$.

- If $A = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & 7 \\ 2 & 8 \\ 3 & 9 \end{bmatrix}$.
- If $B = \begin{bmatrix} 2\pi \sqrt{7} \end{bmatrix}$, then $B^T = \begin{bmatrix} 2\pi \sqrt{7} \end{bmatrix}$.
- A square matrix can be equal to its own transpose. For that to happen, it must be symmetrical with respect to its main diagonal (the diagonal going from the top left to the bottom right), like for example
  \[ \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 8 \\ 4 & 7 & 9 \\ & & \end{bmatrix} \] . Note how the rows and columns are the same.

PROPERTIES (transpose)

See Theorem 1.2 on page 7.

Let $A, B$ be matrices of the same dimensions, let $r$ be a scalar. Then:

- Linearity: $(A + B)^T = A^T + B^T$ and $(rA)^T = rA^T$
- Idempotence: $(A^T)^T = A$
An $m$-dimensional vector is an $m \times 1$ matrix. It is a vertical array of numbers. Vectors are denoted by boldface lowercase letters ($u, v, w, \ldots$) or letters with an arrow on top on the blackboard ($\vec{u}, \vec{v}, \vec{w}, \ldots$).

Elements or entries: The $i$-th entry of a vector $u$ (denoted $u_i$) is the $i$-th number in it. Sometimes we say column vector to emphasize that it has only 1 column. On occasion we’ll need to consider row vectors too ($1 \times n$ matrices).

To save vertical space, we can specify the entries of a (column) vector by writing them in a row and adding the “T” symbol to transpose it, so it becomes a column vector. We can also separate its entries with commas to make it better to read.

Euclidean space: The $m$-dimensional Euclidean space (denoted $\mathbb{R}^m$) is the set of all $m$-dimensional vectors. $\mathbb{R}^1$, $\mathbb{R}^2$ and $\mathbb{R}^3$ are also called the line, the plane and the space and can be visualized geometrically.

- The 4D (4-dimensional) vector $u = \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}$ has elements $u_1 = 5$, $u_2 = 6$, $u_3 = 7$, $u_4 = 8$ and can also be written as $[5, 6, 7, 8]^T$ or as $\begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}^T$.
- The matrix $v = \begin{bmatrix} -2 & 0 & 5 \\ \end{bmatrix}$ is a 3D row vector, but not a column vector.

DEFINITION (vector arithmetic)

Scalar multiplication: Vectors can be multiplied by scalars just like matrices.
Vector addition: Vectors of the same dimension can be added just like matrices.
Zero vector: The zero vector of a given dimension (denoted 0, which is different from the scalar 0) is the vector having all entries equal to zero.

- For $u = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $v = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}$, we have $2u - 3v = \begin{bmatrix} 2 \\ 7 \\ 9 \end{bmatrix}$.

DEFINITION (standard vectors)

The $m$-dimensional standard vectors (denoted $e_1, e_2, \ldots, e_m$) are the $m$ distinct vectors of that dimension that each have only one nonzero entry, equal to 1.

- The standard vectors of $\mathbb{R}^2$ are $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
- The standard vectors of $\mathbb{R}^3$ are $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.
- The rule defining the entries of the standard vector $u = e_k$ is:

$$ u_i = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} $$

This says that, for any fixed $k$, the $k$-th standard vector only has a 1 in the $k$-th position.

Geometric interpretation of vectors

An $m$-dimensional vector can be thought of as a point in $\mathbb{R}^m$, or as the arrow connecting the origin to that point.

This way we can visualize 2D vectors in the plane and 3D vectors in space.

The standard vectors lie on their respective axes, with their arrow heads one unit away from the origin.

- Draw $[1, 2]$ in 2D
- Draw $[1, 2, 3]$ in 3D
1. MATRICES, VECTORS, SYSTEMS OF LINEAR EQUATIONS

Suggested problems from the book

- True or false: 37-56

1.2. Linear combinations, matrix-vector products, some special matrices

Key concepts: linear combinations, matrix-vector products, identity matrix, diagonal matrices, 2D rotation matrices

**DEFINITION (linear combinations)**

A linear combination of \( k \) vectors \( u_1, u_2, \ldots, u_k \) of the same dimension is any vector \( v \) formed by adding scalar multiples of them:

\[
v = c_1 u_1 + c_2 u_2 + \ldots + c_k u_k
\]

where the \( c_i \)'s (called the coefficients) are scalars.

- Examples of linear combinations of 3 vectors \( u, v, w \) are:
  
  \[
  2u - v + 5w, \quad u + 2w, \quad u + v + w, \quad 4v, \quad u, \quad 0 = 0u + 0v + 0w
  \]

- A linear combination of a set of just one vector \( u \) is simply a multiple \( cu \) of \( u \), like:
  
  \[
  2u, \quad -6u, \quad -u, \quad u, \quad 0 = 0u
  \]

- The vector \( v = [5, 5]^T \) can be written as a linear combination of the vectors \( u_1 = [2, 2]^T, u_2 = [6, 4]^T \) and \( u_3 = [1, -1]^T \) in at least two easy ways:
  
  \[
  v = \frac{5}{2} u_1 \quad \text{or} \quad v = u_2 - u_3
  \]

- The zero vector \( 0 \) is always a linear combination of any set of vectors \( u_1, u_2, \ldots, u_k \) in at least one way:
  
  \[
  0 = 0u_1 + 0u_2 + \cdots + 0u_k
  \]

**Not everything is a linear combination of a given set**

Given vectors \( \{u_1, u_2, \ldots, u_k\} \), some vectors \( v \) are linear combinations of them and some are not. Those vectors which are linear combinations may be expressible as a sum of multiples of the \( u_i \) vectors in a unique way or in more than one way.

- \( v = [1, 2, 3]^T \) is not a linear combination of the vectors

\[
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}, \quad \begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}, \quad \begin{bmatrix}
3 \\
5 \\
0
\end{bmatrix}
\]

This is easy to see, because any linear combination of the \( u_i \)'s must have a 0 third entry:

\[
c_1 u_1 + c_2 u_2 + c_3 u_3 = c_1 \begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix} + c_2 \begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix} + c_3 \begin{bmatrix}
5 \\
3 \\
0
\end{bmatrix} = \begin{bmatrix}
1c_1 + 0c_2 + 5c_3 \\
0c_1 + 1c_2 + 3c_3 \\
0c_1 + 0c_2 + 0c_3
\end{bmatrix} = \begin{bmatrix}
c_1 + 5c_3 \\
c_2 + 3c_3 \\
0
\end{bmatrix}
\]

No matter what the \( c_i \)'s are, this can never be equal to \( v \), which has a third entry equal to 3.
• In many cases it may not be so obvious that a vector is not a linear combination of a given set of vectors:

**PROBLEM.** Show that \( v = [2,3]^T \) is not a linear combination of \( u_1 = [1,2]^T \) and \( u_2 = [3,6]^T \).

• This idea of building a system of equations to try to solve for the coefficients also works to find ways to express a vector as a linear combination of some given set:

**PROBLEM.** Show that \( v = [4,0,-1]^T \) is a linear combination of \( u_1 = [0,2,3]^T \), \( u_2 = [2,0,0]^T \) and \( u_3 = [0,2,4]^T \) in a unique way, and find the coefficients.

**PROBLEM.** Find 3 different ways to express \( v = [6,6]^T \) as a linear combination of \( u_1 = [0,2]^T \), \( u_2 = [2,0]^T \) and \( u_3 = [3,3]^T \).

• Every vector in \( \mathbb{R}^n \) is a linear combination of the standard vectors \( e_1, e_2, \ldots, e_m \) in a unique way: the coefficients must be equal to the vector’s entries. For example,

\[
\begin{bmatrix}
4 \\
\pi^3 \\
0 \\
9
\end{bmatrix}
= 4e_1 + \pi^3e_2 + 0e_3 + 9e_4 = 4 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \pi^3 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + 9 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}
\]

• Geometrically, it can be easy to see when a vector is or is not a linear combination of a given set of vectors depending on where it lies in Euclidean space with respect to them:

(Draw why \([1, 2, 3]\) is not a lin comb of \([1, 0, 0], [0, 1, 0], [5, 3, 0]\))

**DEFINITION (matrix-vector product)**

A matrix \( A_{m \times n} \) can multiply a vector \( u_{n \times 1} \) to produce a vector \((Au)_{m \times 1}\).

The number of columns of the matrix must match the dimension of the vector. The dimension of the resulting vector is the number of rows of the matrix.

The \( i \)-th entry of \( Au \), by definition, is calculated by adding the products of the entries from the \( i \)-th row of \( A \) by the corresponding entries of \( u \):

\[
(Au)_i = a_{i1}u_1 + a_{i2}u_2 + \cdots + a_{in}u_n
\]

Order matters: we cannot write \( uA \).

Examples:

\[
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 \\
2 & 0 & 0 & 3 & 1 \\
1 & 0 & 1 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
8 \\
9 \\
10
\end{bmatrix}
= \begin{bmatrix}
80 \\
61 \\
30
\end{bmatrix}
\text{ and } \begin{bmatrix}
1 & 2 \\
3 & 4 \\
5 & 6 \\
7 & 8
\end{bmatrix}
\begin{bmatrix}
10 \\
20
\end{bmatrix}
= \begin{bmatrix}
110 \\
170 \\
230
\end{bmatrix}
\]

• The vector \( Au \) only has the same dimension as \( u \) when \( A \) is a square matrix:

Examples:

\[
\begin{bmatrix}
1 & 2 \\
3 & 0
\end{bmatrix}
\begin{bmatrix}
5 \\
6
\end{bmatrix}
= \begin{bmatrix}
17 \\
15
\end{bmatrix}
\text{ and } \begin{bmatrix}
1 & 0 & 1 \\
1 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
2 \\
3 \\
4
\end{bmatrix}
= \begin{bmatrix}
6 \\
3 \\
6
\end{bmatrix}
\]

• The vector \( Au \) is a linear combination of the vectors that form the columns of \( A \) (denoted \( a_1, a_2, \ldots, a_n \)). The coefficients of the linear combination are the entries of \( u \):

\[
Au = u_1a_1 + u_2a_2 + \cdots + u_na_n
\]

Example:

\[
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6
\end{bmatrix}
\begin{bmatrix}
7 \\
8 \\
9
\end{bmatrix}
= \begin{bmatrix}
1 \times 7 + 2 \times 8 + 3 \times 9 \\
4 \times 7 + 5 \times 8 + 6 \times 9
\end{bmatrix}
= \begin{bmatrix} 1 \times 7 + 2 \times 8 + 3 \times 9 \\
4 \times 7 + 5 \times 8 + 6 \times 9
\end{bmatrix}
= \begin{bmatrix} 1 \times 7 + 2 \times 8 + 3 \times 9 \\
4 \times 7 + 5 \times 8 + 6 \times 9
\end{bmatrix}
= \begin{bmatrix} 7 \times 1 + 8 \times 2 + 9 \times 3 \\
4 \times 7 + 5 \times 8 + 6 \times 9
\end{bmatrix}
\]
PROPERTIES (matrix-vector product)
See Theorem 1.3 on page 24.

Let \( A, B \) be matrices, let \( u, v \) be vectors with dimensions such that the matrix-vector products below are defined, let \( r \) be a scalar. Then:

- Distributive laws: \( A(u + v) = Au + Av \) and \( (A + B)u = Au + Bu \)
- Identity elements: \( A0 = 0 \) and \( 0u = 0 \)
- Associativity with scalar product: \( A(ru) = (rA)u = r(Au) \)

Motivation for matrix-vector products
Suppose that in a population of 100,000 people there are initially 10,000 vegetarians (V) and 90,000 non-vegetarians (NV), but every month 5% of NV become V, and 2% of V become NV. Then the numbers of V and NV after a month are

\[
V : 98\% \times 10000 + 5\% \times 90000 = 9800 + 4500 = 14300
\]

\[
NV : 2\% \times 10000 + 95\% \times 90000 = 85500 + 200 = 85700
\]

For example, to get the number of NV, add the 2% of V that converted into NV to the 95% of people who were NV and remained.

If we put the initial numbers in a vector \([ 10000 \ 90000 ]^T\), then we realize that the operations above amount to multiplying this vector with a transition matrix composed of the fractions of the population that change groups or remain in the same group per month:

\[
\begin{bmatrix}
.98 & .05 \\
.02 & .95
\end{bmatrix}
\begin{bmatrix}
10000 \\
90000
\end{bmatrix} = \begin{bmatrix}
14300 \\
85700
\end{bmatrix}
\]

If we want to know the numbers after 2 months, we can apply the same process to the updated vector after 1 month:

\[
\begin{bmatrix}
.98 & .05 \\
.02 & .95
\end{bmatrix}
\begin{bmatrix}
14300 \\
85700
\end{bmatrix} = \begin{bmatrix}
18299 \\
81701
\end{bmatrix}
\]

It turns out the numbers are slowly converging to a steady-state. This process is an example of a Markov Chain, a topic that will be studied further down the road.

DEFINITION (some special matrices)

- **Identity matrix**: The \( m \)-dimensional identity matrix is the square matrix \( I_{m \times m} \) that only has nonzero entries on the main diagonal, all equal to one. We can denote it by simply \( I_m \) or \( I \).

  Special property of the identity matrix: \( Iu = u \) for all vectors \( u \) of the corresponding dimension.

- **Diagonal matrix**: Any square matrix that only has nonzero entries on the main diagonal.

  The identity matrix is a special type of diagonal matrix.

  - The identity matrices in the first 3 dimensions:
    \[
    I_{1 \times 1} = [1] \quad I_{2 \times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad I_{3 \times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
    \]

  - Verification of the special property for \( I_{2 \times 2} \):
    \[
    \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \cdot a + 0 \cdot b \\ 0 \cdot a + 1 \cdot b \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}
    \]

  - Examples of diagonal matrices in dimensions 3 and 2:
    \[
    \begin{bmatrix}
    1 & 0 & 0 \\
    0 & 2 & 0 \\
    0 & 0 & 3
    \end{bmatrix} \quad \begin{bmatrix}
    0 & 0 & 0 \\
    0 & 1 & 0 \\
    0 & 0 & 0
    \end{bmatrix} \quad \begin{bmatrix}
    -1 & 0 \\
    0 & \pi \\
    0 & 0
    \end{bmatrix} \quad \begin{bmatrix}
    0 & 0 \\
    0 & 0 \\
    0 & 2
    \end{bmatrix}
    \]
DEFINITION (2D rotation matrix)

2D rotation matrix: A matrix of the following form (where $\theta$ is some angle):

$$ R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} $$

When multiplied with a 2D vector, it yields that vector rotated counterclockwise by an angle $\theta$ (this is proved by trigonometric calculations that are not important for us in this class).

- Examples:
  
  $$ R_{\pi/2} \text{ rotates } [3,1]^T \text{ to } [-1,3]^T \text{ (draw)} $$
  
  $$ R_{\pi/4} = \cdots, \text{ apply to } [3,3]^T \text{ (draw)} $$

- Note that $R_0$ is
  $$ \begin{bmatrix} \cos 0 & -\sin 0 \\ \sin 0 & \cos 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_{2 \times 2} $$
  This makes sense: the matrix that rotates vectors by 0 radians (in other words, leaves them exactly as they are) is the identity matrix.

Suggested problems from the book

- Systematic: 1-44
- True or false: 45-63
- Conceptual: 68-70

1.3. Systems of linear equations

Key concepts: linear systems, vector form of solution, row operations, matrix form of a linear system, RREF

DEFINITION (linear system)

Linear system: A system with $m$ equations and $n$ variables in the form

$$ \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n = b_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n = b_m \end{cases} $$

where $a_{ij}$ and $b_i$ are fixed numbers and $x_1, \ldots, x_n$ are the variables. Because we don’t talk about nonlinear systems in a Linear Algebra class, we will often call them just systems.

Solution: A vector $\mathbf{x}$ containing the values of $x_1, \ldots, x_n$ as its entries.

Inconsistent system: A linear system with no solutions.

Consistent system: A linear system with at least one solution.

- The following systems are linear:
  $$ \begin{cases} 2x_1 + 3x_2 = 0 \\ 4x_1 - 3x_2 = 5 \end{cases} \quad \begin{cases} x_1 + x_2 - x_4 + 8x_5 = 7 \\ 9x_1 + 3x_3 = 2 \\ 3x_1 + x_3 = 0 \end{cases} \quad \begin{cases} x_1 - x_2 = 8 \end{cases} $$

- The following systems are not linear:
  $$ \begin{cases} x_1^2 + x_2 = 8 \\ 2x_2 = 4 \end{cases} \quad \begin{cases} \log(x_2) + x_3 = 7 \\ x_1x_2x_3 = 4 \end{cases} \quad \begin{cases} \sqrt{x_3} - x_1 = 0 \end{cases} $$
- Small systems are easy enough to work with by isolating variables in some equations and substituting into others, or by adding appropriate multiples of the equations to one another in order to eliminate certain variables:

\[
\begin{align*}
    x_1 + x_2 &= 7 \\
    x_1 - x_2 &= 3
\end{align*}
\]

(Solve by adding the two equations or by isolating a variable)

**Vector form of solution to a system**

**Vector form:** A way to write the solution(s) of a consistent linear system using vectors.

**Basic variables:** The \(x_i\)'s whose value is determined by the system in terms of the others.

**Free variables:** The \(x_i\)'s whose value we are free to choose to obtain a solution.

- \(x_1 = -8 + 3x_2 + 5x_3, x_2 \text{ free}, x_3 \text{ free}, x_4 = 4 - x_2 + 5x_3\)
- \(x_1 \text{ free}, x_2 = 2 + x_1, x_3 = 6x_1 + 5x_4, x_4 \text{ free}, x_5 = 10 - 3x_1 - 3x_4\)

**Geometric interpretation of a linear system**

Geometrically, the solution(s) to a system with \(n\) variables are vectors in \(\mathbb{R}^n\). Each equation restricts the set of possible solutions to some object (a line if \(n = 2\), a plane if \(n = 3\)). The solutions are the points at the intersection of all these objects.

The system has:
- No solutions if these subspaces have no common intersection.
- A unique solution if these subspaces intersect at a point.
- Infinitely many solutions if these subspaces intersect in some nontrivial manner.

- \(\begin{align*}
    x_1 + 2x_2 &= 4 \\
    -x_1 + x_2 &= -1
\end{align*}\)
- \(\begin{align*}
    x_1 + 2x_2 &= 4 \\
    2x_1 + 4x_2 &= 8
\end{align*}\)
- \(\begin{align*}
    x_1 + 2x_2 &= 4 \\
    3x_1 + 6x_2 &= -5
\end{align*}\)

(Draw pictures)

**Homogeneous systems**

A linear system having \(0\) as vector of independent terms is always **consistent**, because it always has at least the solution \(x = 0\). Such a system is called **homogeneous**.

\[
\begin{align*}
    x_1 + x_2 &= 0 \\
    3x_1 + 3x_2 &= 0 \\
    x_1 + x_2 + x_3 &= 0
\end{align*}
\]

has solution \([0, 0, 0]^T\), but also others.

**Triangular systems**

A **triangular system** (one in which each equation has fewer variables than the previous ones) is easy to solve by **back-substitution**: solve each equation starting from the last and substituting the known values for variables in each.

\[
\begin{align*}
    2x_1 + 3x_2 + x_3 &= 5 \\
    x_2 - x_3 &= 2 \\
    x_3 &= 2
\end{align*}
\]

(solve) (write in triangular pattern)

**DEFINITION (row reduction and elementary row operations)**

**Row reduction:** The simplification of a linear system transforming it into equivalent systems by the use of elementary row operations.
Elementary row operations: 3 types of operations that can be performed to a system to obtain an equivalent one:

- **Multiply an equation by a factor**
  The symbol for multiplication of equation number \( i \) by factor \( c \) is \( cr_i \rightarrow r_i \).

- **Add a multiple of some equation to some other equation**
  The symbol for adding \( c \) times equation \( i \) to equation \( j \) is \( cr_i + r_j \rightarrow r_j \).

- **Swap the position of two equations**
  The symbol for swapping equations \( i \) and \( j \) is \( r_i \leftrightarrow r_j \).

\[
\begin{align*}
2x_1 + 3x_2 + x_3 &= 5 \\
-4x_1 - 6x_2 + 3x_3 &= 0 \\
x_2 - x_3 &= 2
\end{align*}
\]

(perform \( 2r_1 + r_2 \rightarrow r_2 \), then \( r_2 \leftrightarrow r_3 \), then \( (1/5)r_3 \rightarrow r_3 \))

Now we bring in matrix and vector language to study linear systems:

**DEFINITION (matrix form of a linear system)**

A general linear system can always be written in the matrix form \( Ax = b \), where

\[
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}, \quad
b = \begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_m
\end{bmatrix}
\]

The **coefficients matrix** is the \( m \times n \) matrix \( A \).

The **vector of independent terms** is the \( m \)-dimensional vector \( b \).

The **augmented matrix** of the system is the \( m \times (n+1) \) matrix

\[
[ A | b ] = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\
a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn} & b_m
\end{bmatrix}
\]

Its last column is separated from the rest by a vertical bar. It contains all the information in the system. The process of row reduction is most efficiently carried out directly on this matrix.

\[
\begin{align*}
x_1 + x_3 &= 4 \\
-x_1 + 2x_2 + \pi x_3 &= 0 \\
5x_2 + 5x_3 &= 7 \\
x_1 + 2x_2 + 3x_3 &= 0
\end{align*}
\]

**PROBLEM.** Write the system

\[
\begin{align*}
x_1 + 3x_2 + 10x_3 + 5x_4 &= 32 \\
x_1 + 4x_2 + 13x_3 + 6x_4 &= 41 \\
2x_2 + 6x_3 + 5x_4 &= 24 \\
x_1 + 4x_2 + 13x_3 + 12x_4 &= 53
\end{align*}
\]

in matrix form and perform the following operations on its augmented matrix:

\[
- r_1 + r_2 \rightarrow r_2 , - r_1 + r_4 \rightarrow r_4 , -2r_2 + r_3 \rightarrow r_3 , -r_2 + r_4 \rightarrow r_4 , -2r_3 + r_4 \rightarrow r_4 , \]

\[
\frac{1}{3} r_3 \rightarrow r_3 , -5r_3 + r_1 \rightarrow r_1 , -r_3 + r_2 \rightarrow r_2 , -3r_2 + r_1 \rightarrow r_1
\]

Then solve the system and write the solution in matrix form.
DEFINITION (RREF)

Reduced Row Echelon Form: A matrix is said to be in RREF form if:

- The rows with all zero entries, if any, are the last rows.
- The first nonzero entry in any row (called a pivot) is equal to 1.
- The pivots below any given row are to the right of this row’s own pivot (like in a military echelon).
- The entries above and below the pivots are all zero.

Thus every row that’s not all zeroes contains a pivot, but only some columns, called pivotal columns, contain pivots; these columns must then be equal to distinct standard vectors of \( \mathbb{R}^m \).

\[
\begin{bmatrix}
1 & 0 & * & 0 & 0 & 0
\end{bmatrix}
\]

- is in RREF form (indicate the pivotal columns)

\[
\begin{bmatrix}
0 & 1 & * & 0 & * & 0
\end{bmatrix}
\]

- is in RREF form (indicate the pivotal columns)

\[
\begin{bmatrix}
1 & 2 & -1
0 & 0 & 1
\end{bmatrix}
\]

- is not in RREF form (the \(-1\) would have to be 0, because it’s above a pivot)

\[
\begin{bmatrix}
1 & -3 & 0 & 5 & 0 & 0
0 & 1 & 4 & 0 & 0
0 & 2 & 0 & 0 & 1 & 0
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

- is not in RREF form (the 2 would have to be 0, because it’s to the left of a pivot)

\[
\begin{bmatrix}
1 & 2 & 0 & 4 & 4 & 0
0 & 0 & 3 & 5 & 5 & 1
\end{bmatrix}
\]

- is not in RREF form (the 3 would have to be 1, because it’s in a pivot position)

Uniqueness of RREF

See also Appendix E in the book, page 575. A formal proof is difficult to write down carefully because of notation, but the idea is simple.

Given any matrix \( A \), there is only one matrix in RREF form that can be reached from \( A \) by elementary row operations, which we can then call “the RREF form of \( A \)”.

ALGORITHM (nature of solutions of a system based on RREF form)

The best way to study the solutions to a linear system \( Ax = b \) is by reducing its augmented matrix \( [ A | b ] \) to its corresponding RREF form \( [ R | c ] \). Pivotal columns in the \( R \) portion correspond to basic variables, and nonpivotal columns in the \( R \) portion to free variables.

- If the last column \( c \) is pivotal, the system is inconsistent (because the equation corresponding to the row where this pivot is will be \( 0 = 1 \)).
- If all columns are pivotal, except \( c \), the system has a unique solution (because every row corresponds to an equation of the form \( x_i = \cdots \)).
- If there are nonpivotal columns in \( R \), the system has infinitely many solutions (because there will be free variables).

A method for row-reducing the augmented matrix (or any matrix) into its RREF form is given in the next section.
• \[
\begin{bmatrix}
1 & 2 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
corresponds to an inconsistent system

• \[
\begin{bmatrix}
1 & 0 & 0 & 3 \\
0 & 1 & 0 & 4 \\
0 & 0 & 1 & 5
\end{bmatrix}
\]
corresponds to a system with a unique solution

• \[
\begin{bmatrix}
1 & 0 & 2 & 8 \\
0 & 1 & 3 & 9 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]
corresponds to a system with infinitely many solutions

\begin{center}
\textbf{Systems with unique solution must have enough equations}
\end{center}

The RREF analysis shows that a system can only have a unique solution if it has at least as many equations (rows of \(A\)) as variables (columns of \(A\)), because this is the only way for the matrix \(R\) in the RREF \([R \mid c]\) to only have pivotal columns.

But it doesn’t go the other way: there are also inconsistent systems where \(n \geq m\).

An augmented matrix
\[
\begin{bmatrix}
\ast & \ast & \ast & \ast & \ast & \ast \\
\ast & \ast & \ast & \ast & \ast & \ast \\
\ast & \ast & \ast & \ast & \ast & \ast
\end{bmatrix}
\]
corresponds to a system with at least 2 free variables (could be more if there are nonpivotal rows).

\begin{center}
\textbf{Suggested problems from the book}
\end{center}

• Systematic: 1-54
• True or false: 57-76
• Conceptual: 55-56, 79-82

\begin{center}
\textbf{1.4. Gaussian Elimination}
\end{center}

Key concepts: \textit{Gaussian Elimination}, rank and nullity

\begin{center}
\textbf{ALGORITHM (Gaussian Elimination)}
\end{center}

\textbf{Gaussian Elimination} or GE: A systematic method to bring a matrix to RREF form using row operations:

• \textbf{First pass.} Work from top to bottom. For each row:
  \begin{itemize}
  \item Find the pivot position. Look at the rows below to see if any has a pivot occurring before the one you are currently working with, and if so swap these rows.
  \item Create a 1 in the pivot position by scaling the entire row.
  \item Create zeroes below the pivot by adding the appropriate multiple of the row to each row below it.
  \end{itemize}

• \textbf{Second pass.} Work from the bottom up. For each row, create zeroes above its pivot by adding the appropriate multiple of the row to each row above it.

\[
\begin{bmatrix}
0 & 0 & 2 & 10 & -3 & -8 \\
-1 & -5 & 2 & 5 & -5 & -8 \\
2 & 10 & -1 & 5 & 4 & 18 \\
1 & 5 & -1 & 0 & 3 & 8
\end{bmatrix}
\]

\begin{center}
\textbf{PROBLEM.} Find the general solution to the system whose augmented matrix is
\end{center}

\[
\begin{bmatrix}
0 & 3 & 9 & 3 & -6 & -9 & -6 \\
0 & -1 & -2 & 0 & 0 & 1 & 4 \\
0 & 0 & 1 & 1 & -2 & 2 & 14 \\
0 & -2 & -8 & -4 & 8 & 9 & -3
\end{bmatrix}
\]
DEFINITION (rank and nullity)

For any matrix $A$ (doesn’t need to be an augmented matrix), we call:

**Rank**: The number of pivotal columns of its RREF. Notation: $\text{rank}(A)$.

**Nullity**: The number of nonpivotal columns of its RREF. Notation: $\text{nullity}(A)$.

Then $\text{rank}(A) + \text{nullity}(A)$ is always equal to the **number of columns** of $A$.

In the textbook, rank and nullity are defined on page 47 in a different way, but it’s quickly seen on page 48 to be an equivalent definition to ours.

PROBLEM. Find rank and nullity of the following matrices:

$$
\begin{bmatrix}
0 & 0 & 0 & 1 & 2 & 4 \\
0 & 0 & 1 & -1 & -1 & -1 \\
0 & 0 & 1 & 0 & 1 & 3 \\
0 & 0 & 1 & 1 & 3 & 7
\end{bmatrix}
\quad
\begin{bmatrix}
2 & 4 & 2 & 24 \\
2 & 5 & 4 & 33 \\
1 & 2 & 4 & 24
\end{bmatrix}
$$

Possible values for rank and nullity

Let $A$ be an $m \times n$ matrix. Then:

- $\text{rank}(A)$ must be between 0 and the smaller dimension ($m$ or $n$). It’s only 0 if all entries of $A$ are 0.
- $\text{nullity}(A)$ must be between 0 and $n$. It can be larger than $m$.

**Justification.** Neither $\text{rank}(A)$ nor $\text{nullity}(A)$ can be larger than $n$, the number of columns of $A$, by definition. But $\text{rank}(A)$ also cannot be larger than $m$, the number of rows of $A$, because it is equal to the number of pivots, and there can only be at most one pivot per row.

If $A$ has at least one nonzero entry, then the row where this entry is must have a pivot, so there is at least one pivot, hence at least one pivotal column, so $\text{rank}(A) > 0$.

Rank and nullity related to basic and free variables

Given a linear system $Ax = b$, the numbers of basic and free variables are given respectively by the rank and nullity of the coefficients matrix $A$, not of the augmented matrix $[A \mid b]$.

**Justification.** The numbers of basic and free variables are the numbers of pivotal and nonpivotal columns of just the $R$ part of the RREF form $[R \mid c]$ of the augmented matrix. This $R$ part is the RREF of $A$.

Example: $$
\begin{align*}
-x_1 + 3x_2 &= 2 \\
2x_1 + 3x_2 + 4x_3 &= 7 \\
x_1 + 2x_2 + 3x_3 &= 5
\end{align*}
$$

PROBLEM. Consider the system below, which has two parameters $\alpha$ and $\beta$:

$$
\begin{align*}
-x_1 + 3x_2 &= \alpha \\
4x_1 + \beta x_2 &= -8
\end{align*}
$$

For which values of $\alpha$ and $\beta$ does the system have:

a) no solutions?

b) a unique solution?

c) infinitely many solutions?

Suggested problems from the book

- Systematic: 1-42
- True or false: 53-72
- Conceptual: 47-52, 73-84, 87-90
1.6. SPAN

1.5. Applications of linear systems

Section skipped!

1.6. Span

Key concepts: span, generating set

**DEFINITION (span)**

The span of a set of vectors \( \mathbf{u}_1, \ldots, \mathbf{u}_k \) all of the same dimension is the set of all possible linear combinations that can be formed from them:

\[
\text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} = \{c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \cdots + c_k \mathbf{u}_k \mid c_1, c_2, \ldots, c_k \in \mathbb{R}\}
\]

It is a subset of the same \( \mathbb{R}^m \) where each \( \mathbf{u}_i \) belongs, and usually contains infinitely many vectors.

- \( \text{span}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \) (draw)
- \( \text{span}\begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \)
  - Similarly to the item above, the span of the set of all standard vectors in \( \mathbb{R}^m \) is the entire \( \mathbb{R}^m \).
- \( \text{span}\begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \) (draw)
- Span of 0
- \( \text{span}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \)

**ALGORITHM (deciding if a vector is spanned by a given set)**

Suppose we are given vectors \( \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \) and need to decide if a vector \( \mathbf{v} \) belongs to their span. We can form the system \( A\mathbf{x} = \mathbf{v} \) where \( A \) has the \( \mathbf{u}_i \)'s as columns:

\[
A = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{bmatrix} \quad \text{(this is a matrix)}
\]

Then \( \mathbf{v} \) is in \( \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} \) if and only if the system is consistent. If it is, the solutions of the system give the coefficients of the linear combination of \( \mathbf{v} \) in terms of the \( \mathbf{u}_i \)'s.

**PROBLEM.** Decide whether the vectors \([1, 4, 4]^T\) and \([1, 4, 3]^T\) are in the following span:

\[
\text{span}\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}
\]

**PROBLEM.** For which value of \( x \) is the vector \([x, 1, 2]^T\) in the following span?

\[
\text{span}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix}
\]
Justification for the above algorithm. A vector \( \mathbf{v} \) is in \( \text{span}\{ \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \} \) if and only if there exists a way to write it as linear combination of the \( \mathbf{u}_i \)'s, that is, there exist scalars \( c_1, c_2, \ldots, c_k \) such that
\[
c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \cdots + c_k \mathbf{u}_k = \mathbf{v}
\]
The left side can be written as an \( m \times k \) matrix formed by the \( \mathbf{u}_i \)'s as columns (where \( m \) is their dimension), multiplied by the \( k \)-dimensional vector containing the \( c_i \)'s:
\[
\begin{bmatrix}
\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k
\end{bmatrix}
\begin{bmatrix}
c_1 \\
c_2 \\
\vdots \\
c_k
\end{bmatrix}
= \mathbf{v}
\]
This is now saying that \( \begin{bmatrix} c_1 & c_2 & \cdots & c_k \end{bmatrix}^T \) is a solution to the system
\[
\begin{bmatrix}
\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k
\end{bmatrix} \mathbf{x} = \mathbf{v}
\]
So \( \mathbf{v} \) is in the span if and only if the system \( A \mathbf{x} = \mathbf{v} \), where \( A \) has the \( \mathbf{u}_i \)'s as columns, is consistent.

**DEFINITION (generating set for Euclidean space)**

**Generating set for \( \mathbb{R}^m \):** A set of \( m \)-dimensional vectors \( \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \) whose span is the entire \( \mathbb{R}^m \).

- In \( \mathbb{R}^m \), the set \( \{ \mathbf{e}_1, \ldots, \mathbf{e}_m \} \) generates the entire space.
- As we’ve discovered above, the set \( \left\{ \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\} \) generates \( \mathbb{R}^2 \).

A set generates \( \mathbb{R}^m \) if the rank of the corresponding matrix is \( m \)
See also Theorem 1.6 on page 70.

We can decide if \( \{ \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \} \subseteq \mathbb{R}^m \) is a generating set for \( \mathbb{R}^m \) by looking at the RREF \( R \) of the matrix \( A_{m \times k} \) that has these vectors as columns:
- If \( R \) has rows that are completely zero, it’s not a generating set for \( \mathbb{R}^m \).
- If every row of \( R \) has a pivot, it is a generating set for \( \mathbb{R}^m \).

In particular, a generating set for \( \mathbb{R}^m \) must have at least \( m \) vectors (to have one pivot in every row, we need \( m \) pivots, so at least \( m \) columns).

**Justification.** When \( R \) has zero rows, the system \( A \mathbf{x} = \mathbf{v} \) will be inconsistent for some choices of \( \mathbf{v} \) (those that lead to a nonzero last column once the augmented matrix is reduced to RREF form). When \( R \) has no zero rows, every row has a pivot, so any system \( A \mathbf{x} = \mathbf{v} \) will be consistent.

- Do the vectors \( \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \end{bmatrix} \) span \( \mathbb{R}^2 \)?
- Do the vectors \( \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \) span \( \mathbb{R}^3 \)?

**Solutions of a homogeneous system are a span**

When we write the vector form of the solution to a homogeneous system, it naturally comes in the form of the general linear combination of some vectors, because the vector of independent terms is \( \mathbf{0} \):
\[
\mathbf{x} = \mathbf{0} + x_a \mathbf{u} + x_b \mathbf{v} + x_c \mathbf{w} + \cdots
\]
(where \( x_a, x_b, x_c, \ldots \) are the free variables, which we can think of as scalars). Therefore the set of all solutions is the span of the vectors \( \mathbf{u}, \mathbf{v}, \mathbf{w}, \ldots \).

\[
\begin{align*}
x_1 + 2x_2 + 3x_4 &= 0 \\
x_3 + 4x_4 &= 0 \\
0 &= 0
\end{align*}
\]
1.7. Linear dependence and independence

Key concepts: *linearly dependent and independent vectors*

### DEFINITION (linear dependence)

Vectors $u_1, u_2, \ldots, u_k$ are called **linearly dependent** if it’s possible to make a linear combination of them that’s equal to $0$ in a nontrivial way, that is, without all coefficients being equal to $0$:

There exist $c_1, c_2, \ldots, c_k$ not all equal to zero such that $c_1 u_1 + c_2 u_2 + \cdots + c_k u_k = 0$

Otherwise, they are called **linearly independent**:

- If $c_1 u_1 + c_2 u_2 + \cdots + c_k u_k = 0$, then $c_1 = c_2 = \cdots = c_k = 0$

What is dependent or independent is the **set** of the vectors all together, not each individual vector. You can’t say that in a set $\{u_1, u_2, \ldots, u_k\}$ the vector $u_3$ is linearly independent, for example.

- $\{[1, 2, 3, 4]^T, [0, 0, 2, 0]^T, [2, 4, 6, 8]^T\}$ is dependent (trivial combination).
- Any set containing $0$ is dependent. For example, $\{0, u, v, w\}$ is dependent because there is a combination producing $0$ and not using all coefficients equal to $0$:

  $$10 + 0u + 0v + 0w = 0$$

- $\{[1, 0]^T, [0, 1]^T\}$ is independent because the only coefficients that satisfy

  $$c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

  are $c_1 = c_2 = 0$.

- A set with two vectors is dependent if and only if one is a multiple of the other. For example, $\{u, ru\}$ is dependent because we can find a combination of them that equals $0$ and uses at least one nonzero coefficient:

  $$r \cdot u + (-1) \cdot (ru) = 0$$

  (the coefficients used were $r$ and $-1$).
• A set with any number of vectors is dependent if and only if some vector in it is a combination of the others. For example, \{u_1, u_2, 3u_1 - 2u_2\}.

**ALGORITHM (deciding linear dependence using rank)**

Given a set \{u_1, u_2, \ldots, u_k\} containing \(k\) vectors (in some Euclidean space \(\mathbb{R}^m\)), we can decide if it’s dependent or independent based on the matrix \(A\) having the \(u_i\)’s as columns:

\[
A = \begin{bmatrix} u_1 & u_2 & \cdots & u_k \end{bmatrix}
\]

(this is a matrix)

The criterion is as follows:

- If \(\text{rank}(A) = k\), the set is independent.
- If \(\text{rank}(A) < k\), the set is dependent.

**Justification.** Solving the system \(Ax = 0\) means looking for coefficients \(x_1, x_2, \ldots, x_k\) that can be used to write the vector of independent terms (0) as a linear combination of the columns of \(A\) (the vectors \(u_i\)). Therefore:

- If 0 is the only solution, the set is independent.
- If there are multiple solutions, the set is dependent.

This linear system is always consistent because it is homogeneous. What decides if it has a unique or infinitely many solutions is the number of pivotal columns (the rank) of \(A\).

**PROBLEM.** Decide if the following sets are dependent or independent:

\[
\begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & 8 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 & 10 \\ 0 & 4 & 16 \end{bmatrix}
\]

\[
\begin{bmatrix} 1 & 3 & 1 \\ 2 & 1 & 2 \\ 0 & 1 & 3 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 & 1 \\ 2 & 1 & 2 \\ 0 & 1 & 3 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 \end{bmatrix}
\]

By definition, linear independence of a set \(\{u_1, u_2, \ldots, u_k\}\) means that, if we form the system \(Ax = 0\) with \(A\) having the \(u_i\)’s as columns, there will be only one solution. The following theorem says that we can think essentially in the same way replacing 0 with any vector:

**THEOREM (linear dependence if and only if system always has at most one solution)**

See Theorem 1.8 on page 78 (a big list of properties that characterize linearly independent sets, but only the one mentioned here is the nontrivial one).

A set \(\{u_1, u_2, \ldots, u_k\} \subseteq \mathbb{R}^m\) is linearly independent if and only if, for any vector \(b \in \mathbb{R}^m\), the system

\[
\begin{bmatrix} u_1 & u_2 & \cdots & u_k \end{bmatrix} x = b
\]

has at most one solution.

**Proof.** Call \(A\) the matrix \(\begin{bmatrix} u_1 & u_2 & \cdots & u_k \end{bmatrix}\). As with any statement involving an “if and only if”, we must show two things:

a) If the set is independent, then the property about the \(b\) vector is true.

b) If the property about the \(b\) vector is true, then the set is independent.

The one that’s immediate from the definition is b):

- Proof of b) If for any \(b\) there’s at most one solution to \(Ax = b\), then in particular for \(b = 0\) there can’t be solutions other than the trivial solution \(x = 0\), so the set is independent.

- Proof of a) Suppose the set is independent and let \(b\) be any given vector in \(\mathbb{R}^m\). We must show that the system \(Ax = b\) has at most one solution. So suppose that \(x_1, x_2\) are any two solutions. This means that \(Ax_1 = b\) and \(Ax_2 = b\). We need to prove that these solutions are actually the same. The trick is to consider one minus the other: the linearity property of matrix-vector multiplication shows us that the vector \(v = x_1 - x_2\) is a solution to \(Ax = 0\):

\[
Av = A(x_1 - x_2) = Ax_1 - Ax_2 = b - b = 0
\]
But because the set is independent, we know that the only solution to $Ax = 0$ is $x = 0$, therefore $v = 0$. Since $v = x_1 - x_2$, this says that $x_1 = x_2$, as we wanted.

We finish this section with some observations about linearly dependent and independent sets that show up in different contexts:

**Useful remarks about linear dependence and independence**

a) A set of vectors in $\mathbb{R}^m$ with more than $m$ vectors must be **linearly dependent**.

b) A set contained in an independent set must also be **linearly independent**.

c) The pivotal columns of an RREF matrix are **linearly independent**, as are the nonzero row vectors.

Each nonpivotal column is a linear combination of the pivotal columns to its left.

d) The vectors that appear in the vector form solution of a homogeneous system are **linearly independent**.

**Justification.**

a) The corresponding matrix $A$ will have more columns than rows, so it’s impossible to have a pivot in every column in the RREF form. This means there will be free variables in the system $Ax = 0$, that is, infinitely many solutions.

b) If there is no combination of the vectors in a set producing 0 other than that with all coefficients being 0, then the same is true of any subset of it.

c) This is best illustrated with an example. Consider the following RREF matrix:

$$
\begin{bmatrix}
1 & 0 & 5 & 10 & 0 & 1 \\
0 & 1 & -4 & 7 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 3 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

The pivotal columns are distinct standard vectors of $\mathbb{R}^4$, which are always independent. The nonzero rows each have a 1 in a place where all the others have a 0 (the pivots), so that, if we form any linear combination of them, the only way to get a 0 in one of these positions is if the coefficient of that row vector is 0:

\[
c_1 \begin{bmatrix} 1 \\ 0 \\ 10 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ -4 \\ 7 \\ 0 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\]

The first, second and fifth entries imply $c_1 = 0$, $c_2 = 0$ and $c_3 = 0$, so we have independence. To write a nonpivotal column as a combination of the pivotal columns to its left, use its own entries as coefficients:

\[
\begin{bmatrix} 5 \\ -4 \\ 0 \\ 0 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} - 4 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\]

d) If we write the general solution to a homogeneous system $Ax = 0$ in vector form:

\[
x = 0 + x_a u + x_b v + x_c w + \cdots
\]

(where $x_a, x_b, x_c, \ldots$ are the free variables and $u, v, w, \ldots$ are vectors), the vectors that span the set of solutions are $u, v, w, \ldots$, and each has a 1 in the slot corresponding to its free variable and 0’s in the slots corresponding to the other free variables. Just like in item (c), this means they are independent.

**REMEMBER!**

*(See table on page 83.)*

The columns of $A_{m \times n}$ generate $\mathbb{R}^n$ $\iff$ rank($A$) = $m$

The columns of $A_{m \times n}$ are linearly independent $\iff$ rank($A$) = $n$
Suggested problems from the book

- Systematic: 1-62
- True or false: 63-82
- Conceptual: 83-85
Matrices and Linear Transformations

Matrices sometimes can be multiplied together to produce new matrices. The reduction of a matrix into RREF form can be accomplished by multiplying it by some elementary matrices. Some matrices are invertible with respect to this matrix product, and this concept can be used to solve systems as well. A similar algorithm to Gaussian Elimination can compute an inverse if it exists.

But the most important topic in this chapter is by the end: the realization that a matrix is nothing other than an operation that transforms vectors into other vectors satisfying some nice properties, which we can call a linear transformation. Then the concepts just learned of matrix multiplication and inverse are related to composition and invertibility of such transformations as functions.

2.1. Matrix multiplication

Key concepts: matrix-matrix multiplication

**DEFINITION (matrix multiplication)**

A matrix $A$ can be multiplied by a matrix $B$ as long as the number of columns of $A$ is equal to the number of rows of $B$. The result $AB$ is a matrix having as many rows as $A$ and as many columns as $B$:

$$A_{m \times n} B_{n \times k} = (AB)_{m \times k}$$

By definition, the $(i,j)$-th element of the matrix $C = AB$ is obtained by adding the products of the entries in the $i$-th row of $A$ with the corresponding entries in the $j$-th column of $B$:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

The order matters: we cannot write $BA$ for this product.

- $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 0 & 15 \\ 5 & 20 \\ 10 & 25 \end{bmatrix}$
- $\begin{bmatrix} 1 & 2 & 3 & 4 \\ \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 1 & 0 \\ 0 & 4 \\ 2 & 0 \end{bmatrix}$
- $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$

- Sometimes both $AB$ and $BA$ are defined, but they don’t have to be equal: $\begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

- Matrix-vector multiplication is just a special case of matrix-matrix multiplication (a case in which the second matrix has only one column): $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

- $\begin{bmatrix} * & * & * \\ * & * & * \end{bmatrix} \begin{bmatrix} * & * & * \end{bmatrix}$ is undefined.
Motivation for matrix-matrix products
The notion of matrix multiplication comes from iterated matrix-vector multiplication. Suppose we have two matrices $A, B$ and a vector $u$, with small enough dimensions to facilitate the notation:

\[
A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}, \quad u = \begin{bmatrix} x \\ y \end{bmatrix}
\]

Suppose we want to first form the 2D vector $v = Bu$, and then the 2D vector $Av = A(Bu)$ (in general this is only possible if the number $n_A$ of columns of $A$ matches the number $m_B$ rows of $B$, since $v = Bu$ will have dimension $m_B$, while $A$ can only multiply vectors of dimension $n_A$):

\[
Bu = \begin{bmatrix} e & f \\ g & h \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ex + fy \\ gx + hy \end{bmatrix},
\]

\[
A(Bu) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} ex + fy \\ gx + hy \end{bmatrix} = \begin{bmatrix} a(ex + fy) + b(gx + hy) \\ c(ex + fy) + d(gx + hy) \end{bmatrix}
\]

Factoring $x, y$ in these entries, we realize that we can write the end result as just one matrix multiplied with the original vector $u$:

\[
\begin{bmatrix} a(ex + fy) + b(gx + hy) \\ c(ex + fy) + d(gx + hy) \end{bmatrix} = \begin{bmatrix} (ae + bg)x + (af + bh)y \\ (ce + dg)x + (cf + dh)y \end{bmatrix} = \begin{bmatrix} ac + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}
\]

This matrix is what we call $AB$, and we see that its entries are formed by adding the entries of the corresponding row in $A$ multiplied by the ones in the corresponding column in $B$. So we have constructed from $A$ and $B$ a matrix $AB$ with the property that multiplying any vector with it is the same as multiplying this vector with $B$ first and then $A$:

\[
A(Bu) = (AB)u
\]

PROPERTIES (matrix multiplication)
See Theorem 2.1 on page 100 in the book.

Let $A, B, C$ be matrices with dimensions such that the products below are defined. Then:

- Distributive laws: $(A + B)C = AC + BC$ and $A(B + C) = AB + AC$ (order of the products is important)
- Associative law: $A(BC) = (AB)C$
- Transpose law: $(AB)^T = B^T A^T$ (note the order). More generally, $(A_1 A_2 \cdots A_{k-1} A_k)^T = A_k^T A_{k-1}^T \cdots A_2^T A_1^T$ (transpose of product is product of transposes in reverse order)
- Commutative law does not hold: $AB \neq BA$ in general
- The defining property of matrix multiplication: $A(Bu) = (AB)u$ for any vector $u$ of the correct dimension

Proof of the transpose law. To prove a statement that says that certain two matrix expressions are the same, we must show that the general $(i,j)$-th element of one is equal to the one of the other. We assume $A_{m \times n}$ and $B_{n \times k}$ and we call $C = (AB)^T$ and $D = B^T A^T$. For any given $i, j$, the $(i,j)$-th element of $C$ is the $(j,i)$-th element of $AB$, since $C$ is $AB$ transposed. According to the rule for constructing $AB$, this is:

\[
c_{ij} = (AB)_{ji} = a_{j1}b_{1i} + a_{j2}b_{2i} + \cdots + a_{jn}b_{ni}
\]
On the other hand, the \((i, j)\)-th element of \(D\) is obtained by multiplying the matrices \(B^T\) and \(A^T\) in that order, and we can then write the elements of these matrices in terms of the elements of \(A\) and \(B\) by flipping the indices:

\[
d_{ij} = (B^T A^T)_{ij} = b_{i1}^T a_{1j} + b_{i2}^T a_{2j} + \cdots + b_{in}^T a_{nj} = b_{1i}a_{j1} + b_{2i}a_{j2} + \cdots + b_{ni}a_{jn}
\]

Comparing the two equations, we see that the right-hand sides are the same, just with each summand written in a different order (which doesn’t matter for numbers).

Multiplication by diagonal matrices and the identity matrix

Multiplying a matrix \(A\) by a diagonal matrix \(D\) on the left results in a matrix that has the rows of \(A\) scaled by the elements of \(D\).

Multiplying a matrix \(A\) by a diagonal matrix \(D\) on the right results in a matrix that has the columns of \(A\) scaled by the elements of \(D\).

In particular, the identity matrix \(I_{m \times m}\) has the following properties:

\[
IA = A \quad \text{for all } A_{m \times n} \\
AI = A \quad \text{for all } A_{k \times m} \\
AI = IA = A \quad \text{for all } A_{m \times m}
\]

Columns of matrix-matrix product and matrix-vector products

Sometimes it’s useful to remember that the columns of \(AB\) are obtained by the matrix-vector product of \(A\) with each column of \(B\):

\[
AB = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_k \end{bmatrix}
\]

\[
\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 2 & 0 & -5 \\ 0 & 1 & -4 \end{bmatrix} = \begin{bmatrix} 1 \times 2 + 2 \times 0 & 1 \times 0 + 2 \times 1 & 1 \times (-5) + 2 \times (-4) \\ 3 \times 2 + 4 \times 0 & 3 \times 0 + 4 \times 1 & 3 \times (-5) + 4 \times (-4) \\ 5 \times 2 + 6 \times 0 & 5 \times 0 + 6 \times 1 & 5 \times (-5) + 6 \times (-4) \end{bmatrix}
\]

= \[
\begin{bmatrix} 1 \times 2 & 1 \times 0 & 1 \times (-5) \\ 3 \times 4 & 3 \times 0 & 3 \times (-5) \\ 5 \times 6 & 5 \times 0 & 5 \times (-5) \end{bmatrix}
\]

Suggested problems from the book

- Systematic: 5-32
- True or false: 33-49
- Conceptual: 1-4, 52-53, 62-64, 67

2.2. Applications of matrix products

Section skipped!
### 2.3. Invertibility and elementary matrices

Key concepts: matrix inverse, elementary matrices, column correspondence principle

**DEFINITION (matrix inverse)**

Let $A$ be a square matrix. Its inverse, if it exists, is the matrix $A^{-1}$ satisfying

$$AA^{-1} = A^{-1}A = I$$

where $I$ is the identity matrix of the same dimensions as $A$. This concept doesn’t apply to non-square matrices.

**Invertible matrix:** A square matrix that admits an inverse.

**Singular matrix:** A square matrix that doesn’t admit an inverse.

Finding the inverse, when it exists, requires a lot of calculations (next section). But if we have a candidate $B$ for the inverse of $A$ and we want to check that it really is, all we need to verify is whether $AB = I$ and $BA = I$.

- Verify that $\begin{bmatrix} 1 & -3 \\ -2 & 7 \end{bmatrix}$ is the inverse of $\begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix}$.
- Verify that $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ is not the inverse of $\begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$.

**PROBLEM.** Show that $\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$ is not invertible.

We can always determine if a matrix is invertible (and find the inverse, if it is) using the idea in the problem above, which reduces the problem to a few linear systems, but this is very inefficient. In practice we never want to do it that way. There is a better algorithm in the next section.

**Uniqueness of inverse**

Given a square matrix $A$, there can be only at most one inverse, that is, at most one matrix $B$ with the property

$$AB = BA = I$$

That’s why we can call $A^{-1}$ “the inverse of $A$” instead of “an inverse of $A$”.

**Justification.** Suppose $B, C$ are any two matrices with this property:

$$AB = BA = I \quad AC = CA = I$$

We must show that $B$ and $C$ are actually the same matrix. This follows from the associativity property of the matrix product:

$$B = BI = B(AC) = (BA)C = IC = C$$

It was just a matter of a clever trick: inserting $I$ after $B$ (which doesn’t change it), writing this $I$ it as $AC$, applying the associative property, then simplifying from there.

**PROPERTIES (matrix inverse)**

See Theorem 2.2 on page 125 in the book.

Let $A, B$ be invertible matrices of the same dimensions. Then:

- Inverse of the inverse: $(A^{-1})^{-1} = A$
- Inverse of the transpose: $(A^T)^{-1} = (A^{-1})^T$
- Inverse of the product: $(AB)^{-1} = B^{-1}A^{-1}$ (note the order). More generally, $(A_1A_2\cdots A_{k-1}A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_2^{-1}A_1^{-1}$ (the inverse of a product is the product of the inverses in the reverse order)

**Proof of the third property.** To show that the inverse of the matrix $AB$ is the matrix $B^{-1}A^{-1}$, we have to show that the product of these two matrices in either order makes $I$. And indeed, this is a simple consequence of the
2.3. INVERTIBILITY AND ELEMENTARY MATRICES

associativity of the product:

\[(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I\]
\[(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I\]

This implies that the inverse of \(AB\) is \(B^{-1}A^{-1}\), since we know there can be only one matrix \(C\) satisfying \((AB)C = C(AB) = I\). Note how we didn’t have to compute the general \((i,j)\) element of both \((AB)^{-1}\) and \(A^{-1}B^{-1}\) to see that they are equal; it was all manipulations of the properties of matrix products. The same idea applies for products of more than 2 matrices.

You should try similar ideas to prove the other items!

**PROBLEM.** Suppose that \(A, B\) are \(2 \times 2\) matrices with inverses

\[A^{-1} = \begin{bmatrix} 2 & -3 \\ 0 & 1 \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\]

Find the following:

\((BA)^{-1}, (B^T)^{-1}, (A^T B)^{-1}\)

The following formula is only introduced in the book in section 3.1, page 200, but it could be useful to us here:

**FORMULA (inverse of a \(2 \times 2\) matrix)**

The inverse of a \(2 \times 2\) matrix

\[A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\]

is given by

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

(If \(ad - bc = 0\), then it is not invertible).

Commit it to memory: the terms in the main diagonal switch places, the other terms get minus signs, and everything is divided by \(ad - bc\).

**Justification.** Try to find an inverse for \(A\) directly: try to find \(x, y, z, w\) such that

\[\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\]

Multiply out the left hand side and compare entries on both sides. You’ll reach a linear system for \(x, z\) and another for \(y, w\). These systems are consistent and yield the above inverse, as long as \(ad - bc \neq 0\). Otherwise, they don’t have solutions.

**PROBLEM.** Find the inverse of \(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\) and then check that you have the correct answer.

**Motivation for matrix inverse**

The inverse \(x^{-1}\) of a number \(x \neq 0\) is the only number satisfying the property

\[xx^{-1} = x^{-1}x = 1\]

It turns out that it can be calculated as \(1/x\). Its existence is what justifies what we do to solve equations when we “bring a number multiplying on one side to the other side dividing”:

\[xy = z \implies y = \frac{z}{x}\]
What we are really doing is simply multiplying both sides of the equation by \( x^{-1} \), which doesn’t alter what the equation says:

\[
xy = z \iff x^{-1}xy = x^{-1}z \iff y = x^{-1}z
\]

Similarly, if we have a matrix equation where one side features a product of matrices, we can move one to the other side as its inverse, as long as that inverse exists, but we need to be careful with which side to write it in:

\[
AB = C \implies B = A^{-1}C
\]

\[
BA = C \implies B = CA^{-1}
\]

What we are really doing is multiplying both sides by \( A^{-1} \) on the left or right, in order to pair it up with \( A \), producing an \( I \), which leaves the rest of that side of the equation intact:

\[
AB = C \iff A^{-1}AB = A^{-1}C \iff IB = A^{-1}C \iff B = A^{-1}C
\]

\[
BA = C \iff BAA^{-1} = CA^{-1} \iff BI = CA^{-1} \iff B = CA^{-1}
\]

### ALGORITHM (solving a system using matrix inverse)

Suppose we have a linear system containing as many variables as equations \((A\text{ is a square matrix})\):

\[
Ax = b
\]

If \( A \) is invertible, multiplying by \( A^{-1} \) on both sides we conclude that the system has a unique solution given by

\[
x = A^{-1}b
\]

### PROBLEM. Use this idea to solve the system:

\[
\left\{ \begin{array}{l}
  x_1 - 2x_2 = 4 \\
  -5x_1 + 9x_2 = 6
\end{array} \right.
\]

### DEFINITION (elementary matrices)

To every elementary row operation \( \mathcal{O} \) performed on a given matrix \( A_{m \times n} \) corresponds a square matrix \( E_{m \times m} \) called an elementary matrix.

By definition it is the matrix obtained from \( I_{m \times m} \) by performing the same operation \( \mathcal{O} \) on it:

\[
I \xrightarrow{\mathcal{O}} E
\]

Its significance is that the matrix obtained by applying \( \mathcal{O} \) to \( A \) can be calculated by multiplying \( A \) by \( E \) on the left:

\[
A \xrightarrow{\mathcal{O}} EA
\]

- The \( 4 \times 4 \) matrix that performs \( \frac{1}{2}r_3 \rightarrow r_3 \) on a \( 4 \times n \) matrix is …
- The \( 3 \times 3 \) matrix that performs \( \frac{1}{2}r_2 + r_1 \rightarrow r_1 \) on a \( 3 \times n \) matrix is …
- The \( 4 \times 4 \) matrix that performs \( r_1 \leftrightarrow r_3 \) on a \( 4 \times n \) matrix is …

(Verify each by applying to general matrices \( \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \) etc.)

### PROBLEM. For the following \( A, B \), find an elementary matrix \( E \) such that \( B = EA \):

\[
A = \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 0 & -3 \\ 1 & 4 \end{bmatrix}
\]

### Inverses of elementary matrices

The elementary matrices are invertible.

The inverse of the elementary matrix associated to an operation is the elementary matrix associated to the inverse operation:
• The inverse of \( cr_i \rightarrow r_i \) is \( \frac{1}{c} r_i \rightarrow r_i \) (to undo the multiplication of a row by \( c \), you divide it by \( c \)).
• The inverse of \( cr_i + r_j \rightarrow r_j \) is \( -cr_i + r_j \rightarrow r_j \) (to undo the addition of a multiple of row \( i \) to row \( j \), you subtract the same multiple of row \( i \) from row \( j \)).
• The inverse of \( r_i \leftrightarrow r_j \) is itself (to undo a swap of two rows, you swap them again).

Verify these properties for the matrices:

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
1 & 0 \\
-3 & 1
\end{bmatrix}.
\]

These ideas culminate in the following procedure, which is useful when we want to prove properties of the RREF form of a matrix (for example, it could be used to prove that every matrix can be row-reduced to one and only one RREF matrix):

**ALGORITHM (transforming into RREF by multiplying by an invertible matrix)**

The result of applying to a matrix \( A \) the elementary row operations with corresponding elementary matrices \( E_1, E_2, \ldots, E_{k-1}, E_k \) in that order is the product of \( A \) by all those matrices on the left, in reverse order:

\[
A \rightarrow E_k E_{k-1} \cdots E_2 E_1 A
\]

The matrix \( P = E_k E_{k-1} \cdots E_2 E_1 \) is invertible, because it is the product of invertible matrices. Therefore, given a matrix \( A \), there is always an invertible matrix \( P \) such that the RREF form \( R \) of \( A \) can be obtained from \( A \) by multiplying it on the left with \( P \):

\[
R = PA
\]

\( P \) is also easy to compute step-by-step, because products involving elementary matrices are just elementary row operations.

**PROBLEM.** Find an invertible matrix \( P_{3 \times 3} \) such that the \( PA \) is the RREF form of \( A \), where \( A \) is the matrix

\[
\begin{bmatrix}
1 & 0 & -1 & 2 \\
0 & 1 & -2 & 0 \\
0 & -4 & 8 & 3
\end{bmatrix}
\]

Then also find \( P^{-1} \).

The next theorem will be used to justify the algorithm given in the next section for inverting matrices:

**THEOREM (RREF of an invertible matrix)**

A square matrix is invertible if and only if its RREF is the identity \( I \).

In that case, the matrix \( P \) that transforms \( A \) into its RREF \( I \) is the inverse: \( P = A^{-1} \).

**Proof.** This is an “if and only if” theorem, so there are two implications to prove. First we need to prove that, if \( A \) is invertible, then its RREF is \( I \); then we need to prove that, if the RREF form of a matrix is \( I \), it must be invertible.

If \( A \) is invertible, we know that any linear system \( Ax = b \) always has a unique solution \( x = A^{-1} b \). If we tried to solve any such system by GE, this means that every column of \( A \) or of its RREF \( R \) is pivotal. This can only happen, for a square RREF matrix, if it is the identity.

Reciprocally, suppose that \( R = I \). We can use the fact just learned that there is always an invertible matrix \( P \) such that \( R = PA \). In our case, \( I = PA \). Since \( P \) is invertible, we can pass it to the left and conclude \( P^{-1} = A \), that is, \( A^{-1} = P \). So \( P \) is the inverse and \( A \) is invertible.

**THEOREM (column correspondence principle)**

Let \( A \) be a matrix with RREF form \( R \). If a certain column of \( R \) is a linear combination of some other columns of \( R \), then that same column of \( A \) is the same linear combination of those same columns of \( A \).
This can be used to find linear relations between the columns of a matrix (or between any vectors; just build a matrix with them as columns), because the linear relations between the columns of an RREF matrix are always immediate. Therefore, for any matrix:

- The pivotal columns are linearly independent.
- The nonpivotal columns are combinations of the pivotal columns to its left.

- Find a subset of the vectors
  \[ [0, 0, 1]^T, [2, 1, 2]^T, [0, 0, 3]^T, [1, 0, 5]^T, [3, 1, 7]^T \]
  that is linearly independent and such that all the other vectors are combinations of them. Find what those combinations are. (This can be used to reduce a generating set in a span).
- Suppose that the RREF form of the matrix
  \[
  A = \begin{bmatrix}
  6 & 0 & * & * & 3 & * & 0 \\
  5 & 0 & * & * & 3 & * & 8 \\
  6 & * & -3 & * & * & 3 & 0 \\
  5 & * & -4 & * & * & 3 & 0
  \end{bmatrix}
  \]
  is the matrix
  \[
  R = \begin{bmatrix}
  1 & 2 & 0 & 0 & -1 & 5 & 0 & 1 \\
  0 & 0 & 1 & 0 & 4 & 3 & 0 & 2 \\
  0 & 0 & 0 & 1 & 0 & 7 & 0 & 3 \\
  0 & 0 & 0 & 0 & 0 & 0 & 1 & 4
  \end{bmatrix}
  \]
  Complete the missing entries in \( A \).

**Proof idea for the column correspondence principle.** When we perform elementary row operations on a matrix (any operations, not just the ones that reduce it to RREF), any existing linear relationship between some of its columns is preserved. For example, both before and after the operations below, the last column of the matrix is twice the first plus the second:

\[
\begin{bmatrix}
  1 & 2 & 4 \\
  0 & 5 & 5 \\
  1 & 1 & 3
\end{bmatrix} \xrightarrow{r_3+r_1 \rightarrow r_1} \xrightarrow{(1/5)r_2 \rightarrow r_2} \begin{bmatrix}
  2 & 3 & 7 \\
  0 & 1 & 1 \\
  1 & 1 & 3
\end{bmatrix}
\]

This is because the same operations are performed on corresponding elements of any two given columns. Therefore the linear relationships between columns of \( A \) are preserved all the way to its RREF \( R \).

A concise proof can be formulated using the fact, learned in this section, that row operations are performed by multiplication by a matrix on the left (see book).

**Suggested problems from the book**
- Systematic: 1-32, 67-82
- True or false: 33-52
- Conceptual: 53, 59-60, 83

---

### 2.4. Matrix inversion

**Key concepts:** matrix inversion algorithm

**ALGORITHM (matrix inversion)**

Suppose we want to find the inverse of \( A_{m \times m} \).

- Form an \( m \times (2m) \) matrix by appending the identity \( I_{m \times m} \) to the right of \( A \). Separate it with a vertical bar:

\[
[ A \mid I ]
\]
• Perform GE on this augmented matrix to bring \( A \) to its RREF form \( R \) (but \( I \) also changes with every step):
\[
\begin{bmatrix}
A & I
\end{bmatrix} \rightarrow \begin{bmatrix}
R & B
\end{bmatrix}
\]
• If \( R = I \), then \( A \) was invertible and \( B = A^{-1} \).
• If \( R \neq I \), then \( A \) was not invertible.

• Compute \( A^{-1} \) for \( A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 3 & 0 \\ 2 & 5 & 1 \end{bmatrix} \).
• Show that \( \begin{bmatrix} 1 & 0 & 5 \\ 0 & 2 & -2 \\ 0 & 3 & -3 \end{bmatrix} \) is not invertible.
• Solve the system using matrix inverse:
\[
\begin{align*}
2x_1 + 3x_2 &= 6 \\
x_1 + 2x_3 &= 0 \\
2x_1 - 3x_3 &= -2
\end{align*}
\]

**Justification for the inversion algorithm.** The reduction of \( A \) to its RREF \( R \) is achieved by some invertible matrix \( P \):
\[
R = PA
\]
So, when we apply this reduction process not just to \( A \), but to the entire augmented matrix \([A \mid I]\), we are multiplying each of the two parts by \( P \) on the left:
\[
[A \mid I] \rightarrow [PA \mid PI] = [R \mid P]
\]
We already know from the previous section that if \( R \neq I \) then \( A \) is not invertible, and if \( R = I \) then \( A^{-1} \) exists and is equal to \( P \). But \( P \) is precisely what sits on the second part of the reduced augmented matrix. Therefore what sits in the second half of the augmented matrix in the end is the inverse we wanted, provided the first half is \( I \).

Look at theorem 2.6 on page 138 in the book for a list of properties of invertible matrices. You should feel comfortable thinking about why most of them are true, up to item (h). The last 3 are not so trivial.

**Suggested problems from the book**
• Systematic: 1-18, 27-34, 56-63
• True or false: 35-49
• Conceptual: 64-67, 84-87

---

### 2.5. Block multiplication

Section skipped!

---

### 2.6. LU Decomposition

Key concepts: \( LU \) decomposition, solving linear systems with \( LU \) decomposition, algorithm for \( LU \) decomposition

**DEFINITION** (triangular matrices)

- **Main diagonal** of a matrix \( A \): The elements \( a_{ii} \), that is, the ones with equal row and column numbers. These are the elements that start on the upper left corner and go diagonally down from there, irrespective of whether \( A \) is square or not.
- **Upper triangular matrix**: A matrix whose entries below the main diagonal are all 0.
Lower triangular matrix: A matrix whose entries above the main diagonal are all 0.

Triangular matrix: A matrix that is upper or lower triangular.

Unital triangular matrix: A matrix that is upper or lower triangular and has all elements on the main diagonal equal to 1.

- Upper triangular: 
  \[
  \begin{bmatrix}
  1 & 2 & 3 & 4 & 5 \\
  0 & 6 & 7 & 8 & 9 \\
  0 & 0 & \pi & 0 & \sqrt{2}
  \end{bmatrix},
  \begin{bmatrix}
  0 & * & * & * & * \\
  0 & 2 & * & * & * \\
  0 & 0 & 4 & * & * \\
  0 & 0 & 0 & 5 & *
  \end{bmatrix}
  \]

- Lower triangular square unital: 
  \[
  \begin{bmatrix}
  1 & 0 & 0 \\
  2 & 1 & 0 \\
  0 & 1 & 1
  \end{bmatrix},
  \begin{bmatrix}
  1 & 0 & 0 \\
  * & 1 & 0 \\
  * & * & 1
  \end{bmatrix},
  \begin{bmatrix}
  1 & 0 \\
  0 & 1 \\
  * & 1
  \end{bmatrix}
  \]

**DEFINITION (LU decomposition)**

We say that we have an **LU decomposition** of a matrix \(A_{m\times n}\) if we are able to find an upper triangular matrix \(U_{m\times n}\) and a unital lower triangular square matrix \(L_{m\times m}\) such that 

\[A = LU\]

This decomposition doesn’t always exist, and isn’t always unique when it does.

If \(A\) is square and invertible, this decomposition is unique when it exists.

The matrix \(A = \begin{bmatrix} 3 & 2 & 5 \\ 6 & 5 & 14 \end{bmatrix}\) has a possible LU decomposition given by 

\[A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 5 \\ 0 & 1 & 4 \end{bmatrix}\]

A linear system whose coefficients matrix is triangular is easy to solve, so having an LU decomposition for the coefficients matrix \(A\) of a system \(Ax = b\) is useful if we want to solve it:

**ALGORITHM (solving linear systems with LU decomposition)**

Suppose we have a linear system \(Ax = b\) and we know an LU decomposition for \(A\). To solve the system:

- Find the unique solution \(y\) to \(Ly = b\).
- Plug it into the system \(Ux = y\) and solve for \(x\).

Both of these systems are easy to solve by back-substitution because they are triangular.

**Justification.** To solve \(Ax = b\), which is the same as \(L(Ux) = b\) (“\(L\) of something is equal to \(b\)”), first we need to know how to solve \(Ly = b\) (“what is the thing such that \(L\) of it is equal to \(b\)?”). This has a unique solution because all columns of \(L\) are pivotal. Once we have the solution \(y\), we know that that must be the value of \(Ux\) (the “something”) in the equation \(L(Ux) = b\), that is, it must be true that \(Ux = y\). Now this system may be inconsistent or not and it may have infinitely many solutions, since \(U\) is not necessarily square, but nonetheless the back-substitution method easily reveals its nature.

**PROBLEM.** You are told that the following is true:

\[
\begin{bmatrix}
3 & 1 & 1 \\
6 & 3 & -2 \\
3 & 2 & -1
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 \\
2 & 1 & 0 \\
1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
3 & 1 & 1 \\
0 & 1 & -4 \\
0 & 0 & 2
\end{bmatrix}
\]

Use this fact to solve the following system:

\[
\begin{align*}
3x_1 + x_2 + x_3 &= 3 \\
6x_1 + 3x_2 - 2x_3 &= 1 \\
3x_1 + 2x_2 - x_3 &= 1
\end{align*}
\]
PROBLEM. Find the general solution of the following linear system using the LU decomposition given as example after the definition of LU decompositions.
\[
\begin{align*}
3x_1 + 2x_2 + 5x_3 &= 11 \\
6x_1 + 5x_2 + 14x_3 &= 29
\end{align*}
\]

ALGORITHM (finding an LU decomposition)
Suppose we want to find an LU decomposition for a matrix \( A \) (if one exists).
- Perform row operations to reduce \( A \) to an upper triangular form \( U \) (these steps are the same as in the beginning of GE to bring \( A \) to RREF form).
- Find the invertible matrix \( P \) that encodes those operations: \( U = PA \).
- If no row interchanges were used, \( P \) will turn out to be a unital lower triangular matrix, so it is the \( L \) that we seek: \( L = P \).
- If row interchanges were used, \( A \) doesn’t have an LU decomposition.

- Find LU decomp for
  \[
  \begin{bmatrix}
  3 & 1 & 1 & 3 \\
  6 & 3 & -2 & 1 \\
  3 & 2 & -1 & 1 \\
  \end{bmatrix}
  \]
- Find LU decomp for
  \[
  \begin{bmatrix}
  1 & -1 & 2 & 1 \\
  2 & -3 & 5 & 4 \\
  -3 & 2 & -4 & 0 \\
  \end{bmatrix}
  \]

REMARK (when no LU decomposition exists)
When we want to solve \( Ax = b \) but \( A \) doesn’t have an LU decomposition, it’s possible to find a permutation of its rows (that is, change the order of the equations in the system) to create a matrix that does have an LU decomposition. This is just a tweak of the algorithm above and doesn’t add much of a challenge, so we won’t talk about it in this class.

Suggested problems from the book
- Systematic: 1-16
- True or false: 33-40
- Conceptual: 42, 44

### 2.7. Linear transformations and matrices

Key concepts: linear transformation, matrices as linear transformations, standard matrix of a linear transformation

The goal in the last two sections of this chapter is to be able to think of matrices as functions between Euclidean spaces. The concept of function, domain, image etc. generalizes to any sets, but we are only going to consider functions between Euclidean spaces. Note how, while our matrices are usually denoted with dimensions \( m \times n \), a function is usually considered to go from \( \mathbb{R}^n \) to \( \mathbb{R}^m \).

DEFINITION (function between Euclidean spaces)
- Function from \( \mathbb{R}^n \) to \( \mathbb{R}^m \): A procedure or formula that assigns to each element of \( \mathbb{R}^n \) some element of \( \mathbb{R}^m \).
- Image of an element: If \( f : \mathbb{R}^n \to \mathbb{R}^m \) denotes the function, then the element in \( \mathbb{R}^m \) associated to \( u \in \mathbb{R}^n \) is denoted \( f(u) \) and called the image of \( u \).
- Domain: The set \( \mathbb{R}^n \) on which the function acts.
- Codomain: The set \( \mathbb{R}^m \) on which the function takes values.
- Kernel: The set of all values whose image is \( 0 \):
  \[
  \text{Ker}(f) = \{ x \in \mathbb{R}^n \mid f(x) = 0 \} \quad \text{(contained in the domain)}
  \]
**Range:** The set of all values actually attained by the function:
\[ \text{Range}(f) = \{ f(x) \mid x \in \mathbb{R}^n \} \] (contained in the codomain)

- The function \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) defined as \( T(u) = R_\theta u \) for some angle \( \theta \), that is,
  \[
  T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos \theta x_1 - \sin \theta x_2 \\ \sin \theta x_1 + \cos \theta x_2 \end{bmatrix}
  \]
  is called a *rotation*. (Draw picture)
- The function \( T : \mathbb{R}^3 \to \mathbb{R}^3 \) defined as
  \[
  T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}
  \]
  is called a *projection*. (Draw picture) (Point out range)
- The function \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) defined as
  \[
  T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2 \\ x_2 \end{bmatrix}
  \]
  is called a *shear transformation*. (Draw picture)
- The function \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) defined as
  \[
  T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_1 \\ x_2 \end{bmatrix}
  \]
  is called a *reflection*. (Draw picture)
- The function \( f : \mathbb{R}^n \to \mathbb{R}^m \) defined as
  \[
  f(u) = 0
  \]
  is called the *zero function*. (Point out range)
- The function \( f : \mathbb{R}^n \to \mathbb{R}^n \) defined as
  \[
  f(u) = u
  \]
  is called the *identity function*.

**DEFINITION (function associated to a matrix)**

A matrix \( A_{m \times n} \) gives rise to a function
\[
T_A : \mathbb{R}^n \to \mathbb{R}^m \quad \text{(note the order!)}
\]
By definition, this function transforms a vector into its multiplication with \( A \):
\[
T_A(u) = Au \quad \text{for all } u \in \mathbb{R}^n
\]
We call this function the *linear transformation associated to the matrix \( A \)*. Note that it must go from \( \mathbb{R}^n \) to \( \mathbb{R}^m \): the matrix-product \( Au \) is only defined if \( u \) has dimension \( n \), and it yields an \( m \)-dimensional vector.

- The linear transformation associated to
  \[
  A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}
  \]
  is \( T : \mathbb{R}^3 \to \mathbb{R}^2 \) defined by
  \[
  T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ 4x_1 + 5x_2 + 6x_3 \end{bmatrix}
  \]
- The linear transformation associated to
  \[
  A = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}
  \]
  is \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) defined by
  \[
  T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ 2x_1 + 3x_2 \end{bmatrix}
  \]
• The linear transformation associated to \( A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 3 \end{bmatrix} \) is \( T : \mathbb{R}^1 \rightarrow \mathbb{R}^4 \) defined by

\[
T \begin{bmatrix} x \\ -x \\ 0 \\ 3x \end{bmatrix}
\]

• The linear transformation associated to the zero matrix \( O_{m \times n} \) is \( T : \mathbb{R}^n \rightarrow \mathbb{R}^m \) defined by

\[
T \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}
\]

• The linear transformation associated to the identity matrix is the identity transformation. For example, if \( A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \), then \( T_A : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) is

\[
T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
\]

**DEFINITION (linear transformation)**

A function \( f : \mathbb{R}^n \rightarrow \mathbb{R}^m \) that satisfies the following linearity properties:

- \( f(u + v) = f(u) + f(v) \) for any vectors \( u, v \in \mathbb{R}^n \).
- \( f(ru) = rf(u) \) for any scalar \( r \) and vector \( u \in \mathbb{R}^m \).

The linear transformation associated to a matrix satisfies these, as can be easily checked.

- Verify that the transformation \( f[x, y]^T = [2x, x + y, x - y]^T \) is linear.
- Verify that the transformation \( f[x, y, z]^T = [0, 0]^T \) is linear.
- Verify that the transformation \( f[x, y, z]^T = [x^2, y + z]^T \) is not linear.
- Verify that the transformation \( f[x, y]^T = [x + y, 3, x]^T \) is not linear.

**Computing linear transformation on linear combinations**

Let \( T \) be a linear transformation. If we already know its value on some vectors \( u_1, u_2, \ldots, u_k \), we can use them to find its value on any linear combination

\[
v = c_1 u_1 + c_2 u_2 + \cdots + c_k u_k
\]

Because of the linear properties of \( T \), we have

\[
T(v) = c_1 T(u_1) + c_2 T(u_2) + \cdots + c_k T(u_k)
\]

**PROBLEM.** Suppose that a linear transformation \( T : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \) is such that

\[
T \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 4 \end{bmatrix} \quad \text{and} \quad T \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
\]

Compute

\[
T \begin{bmatrix} 3 \\ 6 \end{bmatrix} \quad \text{and} \quad T \begin{bmatrix} 1 \\ 5 \end{bmatrix}
\]

It turns out that every linear transformation (every function satisfying the linearity properties) is associated to some matrix, which is very easy to obtain once we know how the function acts on the standard vectors \( e_i \):
DEFINITION (standard matrix of a linear transformation)

The standard matrix of a linear transformation \( T : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is the \( m \times n \) matrix \( A \) whose columns are the images of the standard vectors of \( \mathbb{R}^n \):

\[
A_T = \begin{bmatrix} T(e_1) & T(e_2) & \cdots & T(e_n) \end{bmatrix} \quad \text{(this is a matrix)}
\]

The linear transformation associated to this matrix is itself:

\[
A_T u = T(u) \quad \text{for all } u \in \mathbb{R}^n
\]

- The standard matrix of \( T[x, y]^T = [5x + 3y, 2x, 8x - y]^T \).
- The standard matrix of the identity transformation.
- The standard matrix of a projection transformation.

Therefore, matrices and linear transformations should be thought of as the same kind of objects. This is the most important thing to be taken from this entire class!

REMEmber!

A matrix \( A_{m \times n} \) can be thought of as a linear transformation \( T : \mathbb{R}^n \rightarrow \mathbb{R}^m \) by the rule

\[
T(u) = Au
\]

A linear transformation \( T : \mathbb{R}^n \rightarrow \mathbb{R}^m \) can be thought of as a matrix \( A_{m \times n} \) by the rule

\[
A = \begin{bmatrix} T(e_1) & T(e_2) & \cdots & T(e_n) \end{bmatrix}
\]

Suggested problems from the book

- Systematic: 25-34, 56-71
- True or false: 35-54
- Conceptual: 1-24, 55, 72-80, 89-90, 95

2.8. Composition and invertibility

Key concepts: onto and one-to-one, column space and null space, composition and invertibility of linear transformations, their relation to matrix product and matrix inverse

DEFINITION (onto and/or one-to-one functions)

A function \( f : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is called:

- **Onto** or surjective if every vector of \( \mathbb{R}^m \) is the image of at least one vector of \( \mathbb{R}^n \) through \( f \):
  
  \[
  \text{Range}(f) = \mathbb{R}^m.
  \]

- **One-to-one** or injective if every vector of \( \mathbb{R}^m \) is the image of at most one vector of \( \mathbb{R}^n \) through \( f \):
  
  if \( f(u) = f(v) \) then \( u = v \) (if two vectors have the same image, they must be the same vector).

- **Bijective** if it is both onto and one-to-one.

- Draw pictures of each case.
- A projection is neither onto nor one-to-one.
- The contraction map \( \mathbb{R}^{m+1} \rightarrow \mathbb{R}^m \) is onto but not one-to-one.
- The extension map \( \mathbb{R}^n \rightarrow \mathbb{R}^{n+1} \) is one-to-one but not onto.

Relation between onto/one-to-one, solutions to systems and rank

(See table on page 188 of the book.)

- A linear transformation \( T : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is onto if and only if every vector \( v \in \mathbb{R}^m \) is equal to \( T(x) \) for some vector \( x \in \mathbb{R}^n \). In terms of the standard matrix \( A \) of \( T \), this means that the system \( Ax = v \) must
always have at least one solution, no matter what \( v \) is. This is the same as saying that the columns of \( A \) are a generating set for \( \mathbb{R}^m \), that is, 
\[
\text{rank}(A) = m
\]

- A linear transformation \( T : \mathbb{R}^n \to \mathbb{R}^m \) is one-to-one if and only if every vector \( v \in \mathbb{R}^m \) is equal to \( T(x) \) for at most one vector \( x \in \mathbb{R}^n \). In terms of the standard matrix \( A \) of \( T \), this means that the system \( Ax = v \) must have at most one solution, no matter what \( v \) is. This is the same as saying that the columns of \( A \) are all pivotal, that is, 
\[
\text{rank}(A) = n \quad \text{or equivalently} \quad \text{nullity}(A) = 0
\]
- As is a consequence of the previous two remarks, a linear transformation \( T : \mathbb{R}^n \to \mathbb{R}^m \) is bijective if and only if its standard matrix \( A \) is invertible. In particular \( A \) must be square, that is, \( m = n \). A linear transformation between different Euclidean spaces cannot be bijective.

**PROBLEM.** Determine if the following are onto, one-to-one, both or neither.

\[
T[x, y, z]^T = [2x + z, y] \quad U[x, y]^T = [3y - x, x + y]^T \quad V[x, y, z]^T = [2y, x, 3z]^T
\]

**DEFINITION (column space and null space)**

Let \( A_{m \times n} \) be a matrix.

**Column space:** The span of the columns of \( A \). Denoted \( \text{Col}(A) \).

\[
\text{Col}(A) = \text{span}\{a_1, a_2, \ldots, a_n\} \subseteq \mathbb{R}^m
\]

**Null space:** The set of all the vectors \( u \) such that \( Au = 0 \). Denoted \( \text{Null}(A) \).

\[
\text{Null}(A) = \{u \in \mathbb{R}^n \mid Au = 0\} \subseteq \mathbb{R}^n
\]

Find generating sets for the column space and null space of

\[
\begin{bmatrix}
1 & 2 & 3 \\
1 & 1 & 0
\end{bmatrix}
\]

**Column space and row space related to range and kernel**

If \( A \) is a matrix and \( T_A \) is its related linear transformation, then

\[
\text{Col}(A) = \text{Range}(T_A) \quad \text{Null}(A) = \text{Ker}(T_A)
\]

*Justification.* Remember that \( T_A \) is the linear transformation defined by

\[
T_A(x) = Ax
\]

The column space

\[
\text{Col}(A) = \text{span}\{a_1, a_2, \ldots, a_n\}
\]

is the set of all linear combinations of the columns \( a_i \) of \( A \). We’ve already seen how any such linear combination, with coefficients \( x_1, x_2, \ldots, x_n \), can be written in the form \( Ax \), which is by definition \( T_A(x) \). Therefore \( \text{Col}(A) \) is the set of all vectors that are the image of some vector \( x \) via \( T_A \), that is, \( \text{Range}(T_A) \).

The null space

\[
\text{Null}(A) = \{x \text{ such that } Ax = 0\}
\]

is the set of all vectors \( x \) such that \( T_A(x) = 0 \), that is, \( \text{Ker}(T_A) \).

Find generating sets for the range and kernel of \( T[x, y]^T = [x + 4y, y, x + 3y]^T \).

**THEOREM (onto and one-to-one in terms of size of column space and null space)**

Let \( T : \mathbb{R}^n \to \mathbb{R}^m \) be a linear transformation with standard matrix \( A \). Then:

a) \( T \) is onto if and only if the column space of \( A \) is the largest possible: \( \text{Col}(A) = \mathbb{R}^m \).

b) \( T \) is one-to-one if and only if the null space of \( A \) is the smallest possible: \( \text{Null}(A) = \{0\} \).

*Proof.*
a) This is the trivial part, direct by the definition of range and onto.
b) \( 0 \) is always in Null\((A)\), because \( A0 = 0 \) no matter what \( A \) is. But in order for \( T \) to be one-to-one, we need all systems \( Ax = v \) to have at most one solution, and we already know that this is equivalent to asking that just the system \( Ax = 0 \) have at most one solution, which is to say that \( 0 \) is the only element in Null\((A)\).

**DEFINITION** (composition and inverse of functions)

**Composition**: Whenever we have two functions \( f: \mathbb{R}^n \rightarrow \mathbb{R}^m \) and \( g: \mathbb{R}^m \rightarrow \mathbb{R}^k \) (note that the codomain of \( f \) is the domain of \( g \)), we can form a function \( g \circ f: \mathbb{R}^n \rightarrow \mathbb{R}^k \) (note the order!), called their composition, defined by

\[
(g \circ f)(u) = g(f(u))
\]

This simply consists of first applying \( f \), then \( g \). *Note: The book uses the notation \( gf \) instead of \( g \circ f \).*

**Inverse**: If a function \( f: \mathbb{R}^n \rightarrow \mathbb{R}^m \) is bijective, there exists a function \( f^{-1}: \mathbb{R}^m \rightarrow \mathbb{R}^n \), called its inverse, with the property that

\[
\begin{align*}
  f^{-1}(f(u)) &= u \quad \text{for all } u \in \mathbb{R}^n \\
  f(f^{-1}(v)) &= v \quad \text{for all } v \in \mathbb{R}^m
\end{align*}
\]

The inverse function undoes the transformation that the function did. Note that, in this class, we only consider functions which are linear transformations, and in that case we already know that, in order for \( f \) to be bijective, the domain and codomain must be the same \((m = n)\).

Draw pictures!

- Find the composition in both orders:
  \[
  T[x, y]^T = [x + y, 2x, x - y]^T \\
  U[x, y, z]^T = [3z + 4x, x + y + z]^T
  \]

- Show that
  \[
  T[x, y]^T = [x + y, x]^T \\
  U[x, y]^T = [y, x - y]^T
  \]

are inverses of each other.

The next theorem gives a new meaning for matrix multiplication and inverse in this new language of functions:

**THEOREM** (composition and inverse in terms of matrices)

*Look at pages 186 and 187 in the book for some insights and a proof.*

a) Let \( A_{m \times n} \) and \( B_{n \times k} \) be matrices with corresponding linear transformations \( T_A: \mathbb{R}^n \rightarrow \mathbb{R}^m \) and \( T_B: \mathbb{R}^k \rightarrow \mathbb{R}^n \). Then the composition \( T_A \circ T_B: \mathbb{R}^k \rightarrow \mathbb{R}^m \) is also a linear transformation, and its associated standard matrix is the matrix product \( AB \).

b) Let \( A_{m \times m} \) be an invertible matrix with corresponding linear transformation \( T_A: \mathbb{R}^m \rightarrow \mathbb{R}^m \). Then the inverse transformation \( T_A^{-1}: \mathbb{R}^m \rightarrow \mathbb{R}^m \) is also a linear transformation, and its associated standard matrix is the matrix inverse \( A^{-1} \).

- Find the inverse transformation:
  \[
  T[x, y]^T = [5y - x, 8x - 4y]^T
  \]

- Compose \( T \) with itself:
  \[
  T[x, y, z]^T = [10x, y + z, 3y]^T
  \]

- What is the composition \( R_{30^\circ} \circ R_{60^\circ} \)? Use matrix product to verify it.

- Do the same for \( R_{90^\circ} \circ T \circ R_{90^\circ} \), where \( X = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \) is the reflection around the \( x \)-axis.

**Suggested problems from the book**

- **Systematic**: 1-40, 69-90
- **True or false**: 41-60
- **Conceptual**: 61-68, 91-94
Determinants

Every square matrix has a number associated to it, called its determinant, which is not so easy to compute, but it has the very useful property that it is zero only for those matrices that are not invertible. It also has geometric significance, allowing the computation of areas and volumes, and finds applicability at yet another algorithm that can solve linear systems: the Cramer Rule.

This entire chapter concerns only square matrices.

3.1. Cofactor expansion

Key concepts: determinants of 1 x 1, 2 x 2 and 3 x 3 matrices, cofactor expansion, determinant of triangular matrices, areas and volumes in terms of determinants

Earlier we found a formula for inverting a 2 x 2 matrix \([a\ b\ c\ d]\), but it depended on the number \(ad - bc\) not being zero. It is a general fact that, whenever we have a square matrix of any dimension, there is a number that we can compute with its entries that tells us if it is invertible (it is invertible if and only if this number is not zero):

**DEFINITION (determinant)**

Every square matrix \(A\) has a number associated to it, called its determinant and denoted \(\det(A)\).

If we specify the entries of the matrix, the determinant can also be denoted with vertical bars:

\[
\begin{vmatrix}
  a_{11} & a_{12} & \cdots & a_{1m} \\
  a_{21} & a_{22} & \cdots & a_{2m} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mm}
\end{vmatrix}
\]

Determinant of 1 x 1 matrix: \(|a| = a\) (these bars do not mean absolute value in this context!)

Determinant of 2 x 2 matrix:

\[
\begin{vmatrix}
  a & b \\
  c & d
\end{vmatrix} = ad - bc
\]

Determinant of 3 x 3 matrix:

\[
\begin{vmatrix}
  a & b & c \\
  d & e & f \\
  g & h & i
\end{vmatrix} = aei + bfg + cdh - ceg - afh - bdi
\]

To remember this formula, write copies of the first two columns of the matrix in front of it, then add the product of the descending diagonals and subtract the product of the ascending diagonals:

\[
\begin{vmatrix}
  a & b & c \\
  d & e & f \\
  g & h & i
\end{vmatrix}
\]

Determinants of larger matrices cannot be computed this way. We’ll look at them later.

- Compute the determinants of:

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

\[
\begin{bmatrix}
  1 & 2 & 3 \\
  4 & 5 & 6 \\
  7 & 8 & 9
\end{bmatrix}
\]
Show that, for any $2 \times 2$ matrices $A, B$, we have
\[ \det(AB) = \det(A) \det(B) \]

The reason to consider this strange concept is:

**THEOREM (determinant and invertibility)**

A square matrix $A$ is invertible if and only if $\det A \neq 0$.

- Compute the determinant of a rotation matrix in $\mathbb{R}^2$ and use it to show that it is invertible.
- Determine the values of $x$ for which the matrix
  \[
  \begin{bmatrix}
  17 - x & -6 \\
  45 & -16 - x
  \end{bmatrix}
  \]
  is not invertible.
- Have the class come up with a matrix where the 3rd column is a combination of the first two, then show $\det = 0$.

The algorithm below can be thought of as the definition of the determinant in dimensions larger than 3 (but it works for 1, 2, 3 too). It is a recursive definition, because it requires calculating determinants of one fewer dimension.

**ALGORITHM (cofactor expansion)**

Let $A_{m \times m}$ be a square matrix.

- For any $1 \leq i \leq m$ and $1 \leq j \leq m$, the **cofactor** $c_{ij}$ is defined to be the number given by $(-1)^{i+j}$ times the determinant of the $(m-1) \times (m-1)$ matrix obtained from $A$ by eliminating row $i$ and column $j$.
- To get the factors of $(-1)^{i+j}$ right, it’s useful to draw a copy of the matrix with + and − signs in a checkerboard pattern, starting with + in the top left corner:

\[
\begin{pmatrix}
+ & - & + & \cdots & \pm \\
- & + & - & \cdots & \mp \\
+ & - & + & \cdots & \pm \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\pm & \mp & \pm & \cdots & +
\end{pmatrix}
\]

Note that the signs on the main diagonal are all +.
- The **determinant** $\det A$ is the number calculated as follows: choose a row or column (it doesn’t matter which), compute the product of each entry there with the corresponding cofactor and add it all up:
  \[
  \begin{align*}
  \det A &= a_{11}c_{11} + a_{12}c_{12} + \cdots + a_{im}c_{im} \quad \text{if we choose row } i \\
  \det A &= a_{1j}c_{1j} + a_{2j}c_{2j} + \cdots + a_{mj}c_{mj} \quad \text{if we choose column } j
  \end{align*}
  \]

It doesn’t matter along which row or column we do the cofactor expansion, so it’s better to always choose one that will make the calculations simple (one containing many zeroes).

*The proof of why this process doesn't depend on the row or column and why it yields a number that is zero if and only if $A$ is not invertible is quite complicated (a good practice for mathematical induction, but a notational nightmare). See page 219 if you're interested.*

- Verify that the formulas given for $2 \times 2$ and $3 \times 3$ determinants are compatible with the cofactor expansion.
- The matrix
  \[
  A = \begin{bmatrix}
  0 & 3 & 0 & 0 \\
  1 & 9 & 2 & 1 \\
  1 & 7 & 2 & 2 \\
  1 & 8 & 2 & 3
  \end{bmatrix}
  \]
  is not invertible (for example, because column 3 is a multiple of column 1, so the rank is not maximal). Compute $\det(A)$ to see that it is zero.
**PROBLEM.** Compute the following:

\[
\begin{vmatrix}
  3 & 1 & 0 & -2 \\
  1 & 2 & 1 & 5 \\
  0 & 1 & 2 & 0 \\
  1 & 0 & -1 & -2
\end{vmatrix}
\quad
\begin{vmatrix}
  1 & 0 & -2 & 0 & 4 \\
  2 & 0 & -1 & 2 & -1 \\
  1 & 0 & 1 & 0 & 2 \\
  3 & 3 & 2 & 0 & -1 \\
  1 & 0 & 1 & 0 & 2
\end{vmatrix}
\]

**Zero determinant means linearly dependent columns and rows**

We already know that, for square matrices \( A \), being invertible is equivalent to having independent columns. Therefore, \( \det(A) = 0 \) means that the **columns** of \( A \) are **dependent**.

Because taking the transpose of a matrix transforms its rows into columns and vice-versa, and the algorithm for the determinant can be carried out through a row or a column, we have the following property:

\[ \det(A) = \det(A^T) \]

So it follows that \( \det(A) = 0 \) also means that the **rows** of \( A \) are **dependent**.

This is a remarkable fact, not at all trivial! It means that, whenever we have a square matrix where some column is a linear combination of the other columns, the same must be true of some row with respect to some other rows.

Have the class come up with a matrix where a column is a combination of the other columns. Then use GE on the transpose to find a row that is a combination of the other rows.

**Determinant of triangular matrix**

The determinant of a triangular matrix (therefore also of any diagonal matrix) is the product of the elements on the main diagonal.

In particular, \( \det I_{m \times m} = 1 \).

An example can convince you of this better than a formal proof:

\[
\det \begin{bmatrix}
  1 & * & * & * \\
  0 & 2 & * & * \\
  0 & 0 & 3 & * \\
  0 & 0 & 0 & 4
\end{bmatrix} = \cdots
\]

(Use cofactor expansion on the last row)

**Geometric meaning of determinant**

Let \( \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \in \mathbb{R}^m \). The **absolute value of the determinant** of the matrix having these vectors as columns gives the area, volume, hypervolume etc. of the shape that they form in \( m \)-dimensional space:

\[
\left| \det [ \mathbf{u}_1 \mathbf{u}_2 \cdots \mathbf{u}_m ] \right|
\]

For example:

- The area of the parallelogram whose sides are the arrows corresponding to vectors \([a, b]^T\) and \([x, y]^T\) is

  \[
  A = \left| \det \begin{bmatrix}
  a & x \\
  b & y \\
\end{bmatrix} \right|
  \]

- The volume of the parallelepiped whose sides are the arrows corresponding to vectors \([a, b, c]^T\), \([x, y, z]^T\) and \([p, q, r]^T\) is

  \[
  V = \left| \det \begin{bmatrix}
  a & x & p \\
  b & y & q \\
  c & z & r
\end{bmatrix} \right|
  \]

- Draw the parallelepiped with vertices

  \[(0, 0, 0), (2, 1, 0), (2, 3, 0), (0, 2, 0), (0, 5, 5), (2, 6, 5), (2, 8, 5), (0, 7, 5)\]

  and find its volume.
• Find the area of the triangle with vertices 

\((0,0), (1,2), (3,3)\)

Suggested problems from the book

• Systematic: 1-44, 65
• True or false: 45-52, 54-63
• Conceptual: 66-72

3.2. Properties of determinants

Key concepts: properties of determinants, Cramer’s Rule

PROPERTIES (determinants)
See Theorem 3.4 on page 214.

Let \(A, B\) be square matrices of the same dimension \(m \times m\), let \(r\) be a scalar. Then

* Determinant of the product: \(\text{det}(AB) = \text{det}(A) \text{det}(B)\)
* Determinant of the transpose: \(\text{det}(A^T) = \text{det}(A)\)
* Determinant of the inverse: If \(A\) is invertible, then \(\text{det}(A^{-1}) = \frac{1}{\text{det}(A)}\)
* Homogeneity: \(\text{det}(rA) = r^m \text{det}(A)\) (careful! \(r\) doesn’t come out of the determinant unaffected, but rather raised to the matrix’ dimension!)
* Non-linearity: \(\text{det}(A + B) \neq \text{det}(A) + \text{det}(B)\) in general

• Suppose that \(A, B\) are \(3 \times 3\) matrices with \(\text{det}(A) = 2\) and \(\text{det}(B) = -5\). Let \(C = 4A\) and \(D = B^T\). Find \(\text{det}(CD^2)\).
  • Show that it’s false that \(\text{det}(A + B) = \text{det}(A) \text{det}(B)\) for all square matrices \(A, B\) of the same size.
  • Use the property \(\text{det}(AB) = \text{det}(A) \text{det}(B)\) to prove the formula for the determinant of an inverse matrix.
  • Under what conditions is \(\text{det}(-A) = -\text{det}(A)\)?

REMARK (proving the properties of determinants)

We are not going to worry about the proof of the determinant properties.

In theory it’s possible to prove them using the cofactor expansion algorithm as the definition of determinants, probably by induction on the dimension of the matrix, but this would be terribly complicated and not very illuminating.

The correct approach is to first realize that there is a better, indirect definition of the determinant: it is essentially the only so-called multilinear function of \(m\) \(m\)-dimensional vectors that has a property called total antisymmetry (it translates to the fact that switching two columns in a matrix alters the determinant by a minus sign). One has to prove that there is only one such function, but once that has been established, it becomes easy to prove the properties and show that the cofactor expansion is a way to calculate this function.

Determinant of the elementary matrices

The determinant of the elementary matrices is easy to compute by cofactor expansion:

* If \(E\) corresponds to \(cr_i \rightarrow r_i\), then \(\text{det}(E) = c\).
* If \(E\) corresponds to \(cr_i + r_j \rightarrow r_j\), then \(\text{det}(E) = 1\).
* If \(E\) corresponds to \(r_i \leftrightarrow r_j\), then \(\text{det}(E) = -1\).

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & c & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\quad \text{(diagonal matrix)}
\]
3.2. PROPERTIES OF DETERMINANTS

- \[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & c & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

(see cofactor expansion along the row containing \(c\))

- \[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

(see cofactor expansion along rows that were not exchanged)

- We can use these to conclude that
  - Multiplying a row of a matrix by some \(c\) multiplies its determinant also by \(c\).
  - Adding a multiple of some row of a matrix to another row doesn’t change its determinant.
  - Switching two rows of a matrix multiplies its determinant by \(-1\).

This is because each of these changes is carried out by multiplication with an elementary matrix, and
\[
\det(\mathbf{EA}) = \det(\mathbf{E}) \det(\mathbf{A}).
\]

**ALGORITHM (calculating determinant by reducing to triangular matrix)**

If we want to compute the determinant of some matrix \(\mathbf{A}\), we can proceed as follows:
- Perform row reduction to transform \(\mathbf{A}\) into an upper triangular matrix \(\mathbf{U}\). The determinant of \(\mathbf{U}\) is easy (the product of the diagonal elements).
- Let \(\mathbf{P}\) be the square matrix that accomplishes the reduction: \(\mathbf{U} = \mathbf{PA}\).
- \(\mathbf{P}\) is the product of a few elementary matrices \(\mathbf{E}_k, \mathbf{E}_{k-1}, \ldots, \mathbf{E}_2, \mathbf{E}_1\), so its determinant is easy to find:
  \[
  \det(\mathbf{P}) = \det(\mathbf{E}_k) \det(\mathbf{E}_{k-1}) \cdots \det(\mathbf{E}_2) \det(\mathbf{E}_1)
  \]
- Because \(\mathbf{U} = \mathbf{PA}\), we have \(\det(\mathbf{U}) = \det(\mathbf{P}) \det(\mathbf{A})\). Therefore
  \[
  \det(\mathbf{A}) = \frac{\det(\mathbf{U})}{\det(\mathbf{P})}
  \]

- Apply this to find \(\det\begin{bmatrix} 1 & 2 & 1 \\ 3 & 4 & 8 \\ 1 & 1 & 2 \end{bmatrix}\).

- Apply this to find all values of \(x\) for which \[
\begin{bmatrix}
1 & 2 & -1 \\
2 & 3 & x \\
0 & x & -15
\end{bmatrix}
\]
is not invertible.

Next we give an application of determinants for solving linear systems. The method is actually not very efficient, but has historical significance and is sometimes useful for proving theorems about systems.

**ALGORITHM (Cramer’s Rule)**

Suppose we want to solve a linear system \(\mathbf{Ax} = \mathbf{b}\) where the coefficients matrix \(\mathbf{A}\) is **square** and **invertible**. We know the solution exists and is unique. Cramer’s Rule is the following:
- For each column \(j\) of \(\mathbf{A}\), compute the determinant \(d_j\) of the matrix \(\mathbf{A}\) with column \(j\) replaced by \(\mathbf{b}\).
- Compute the determinant of \(\mathbf{A}\).
- The solution of each variable is
  \[
  x_j = \frac{d_j}{\det(\mathbf{A})}
  \]

- \[
\begin{cases}
3x_1 + 8x_2 = 4 \\
2x_1 + 6x_2 = 2
\end{cases}
\]

- \[
\begin{cases}
x_1 + 2x_2 + 3x_3 = 2 \\
x_1 + x_2 + x_3 = 3 \\
x_1 + x_2 - x_3 = 1
\end{cases}
\]

*Proof for Cramer’s Rule.* Let’s call \(M_j\) the matrix \(\mathbf{A}\) with column \(j\) replaced by \(\mathbf{b}\). Also call \(X_j\) the matrix \(\mathbf{I}\) with column \(j\) replaced by \(\mathbf{x}\), where \(\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}\) is the solution that we seek to compute. Somewhere in this proof we must use the fact that \(\mathbf{Ax} = \mathbf{b}\).
The trick is to note that $\det(X_j) = x_j$ (use a cofactor expansion along row $i$) and that $AX_j = M_j$ (compute each column separately and use $Ax = b$). Then we apply determinants to this last equation:

$$\det(A) \det(X_j) = \det(M_j) \implies \det(A)x_j = \det(M_j) \implies x_j = \frac{\det(M_j)}{\det(A)}$$

**REMARK** (a matrix is invertible if and only if it has a one-sided inverse)

Suppose $A$ is a square matrix. To prove that some matrix $B$ is its inverse, we learned that we have to verify two things:

$$AB = I \quad \text{and} \quad BA = I$$

But because of the properties of determinants, it turns out to be enough to check just one of them!

Suppose for example that we only checked $AB = I$. Taking determinants, we have

$$\det(A) \det(B) = \det(I) = 1$$

In particular, none of the numbers $\det(A)$ and $\det(B)$ can be zero, so both $A$ and $B$ must be invertible. But then multiplying both sides of $AB = I$ on the left by $A^{-1}$, we conclude $B = A^{-1}$, so $B$ is the inverse of $A$.

---

Suggested problems from the book

- **Systematic**: 1-38, 59-66
- **True or false**: 39-57
- **Conceptual**: 67-75, 77
CHAPTER 4

Linear subspaces

The kinds of sets that can be spanned by vectors are characterized by so-called linearity properties and receive the name of linear subspaces or vector subspaces. Examples are points, lines and planes (which we can imagine), and 3D, 4D, 5D etc. subspaces contained inside even larger spaces (which we can only treat analytically). The dimension of a subspace is “the number before D”: a measure of how many independent vectors must be used to span it. Any such set of generating vectors is called a basis for it, and there are infinitely many bases for any given subspace (other than a single point), just like a 2D sheet of paper can be generated by any two vectors drawn in it that are not parallel.

Matrices and the linear transformations related to them also have subspaces associated to them: the column space and row space (spanned by its columns and rows) and the null space (already introduced). We’ll learn how to find bases for them. Understanding what these subspaces are gives us a complete characterization of anything we’d like to know about a matrix: if it came from a linear system, they will tell us rank, nullity, solution formulas etc.; if it came from a linear transformation, they will let us rewrite it in simpler forms by using a change of variables.

4.1. Linear subspaces

Key concepts: subspace, generating sets for column space and null space

Euclidean space \( \mathbb{R}^m \) is a set whose elements can be added together and multiplied by scalars to produce other elements that live in that same set. It also includes a very special element \( 0 \) that is the identity element of this sum operation. These properties are the essence of all Euclidean spaces.

But some subsets of \( \mathbb{R}^m \) also have these properties, which makes them look like Euclidean spaces \( \mathbb{R}^n \) of a smaller dimension. For example, a line going through the origin inside of \( \mathbb{R}^2 \) looks like \( \mathbb{R}^1 \), and a plane going through the origin inside of \( \mathbb{R}^3 \) looks like \( \mathbb{R}^2 \). These subsets receive a special name:

**DEFINITION (subspace)**

A **vector subspace**, linear subspace or simply **subspace** of \( \mathbb{R}^m \) is a (nonempty) subset \( W \subseteq \mathbb{R}^m \) that satisfies the following properties:

- \( 0 \in W \)
- If \( u, v \) are any two elements of \( W \), then their sum \( u + v \) is also in \( W \)
- If \( u \) is any element of \( W \) and \( r \) is any scalar, then \( ru \) is also in \( W \)

The **trivial subspaces** of \( \mathbb{R}^m \) are the largest and smallest possible: \( \mathbb{R}^m \) itself and \( \{0\} \). They satisfy the properties above, so they can be called subspaces.

- \( \{ [x, y, z]^T \mid x + 2y + 3z = 0 \} \) is a subspace of \( \mathbb{R}^3 \).
- \( \{ [x, y]^T \mid y = x^2 \} \) is **not** a subspace of \( \mathbb{R}^2 \). (Draw picture).
- \( \{ [x, y]^T \mid y = 3x \} \) is a subspace of \( \mathbb{R}^2 \). (Draw picture).
- \( \{ [x, y]^T \mid y = 2x + 3 \} \) is **not** a subspace of \( \mathbb{R}^2 \). (Draw picture).
- \( \{ [x, y]^T \mid x, y \geq 0 \} \) is **not** a subspace of \( \mathbb{R}^2 \). (Draw picture).
- If \( u \in \mathbb{R}^m \), the set of all multiples of \( u \) is a subspace.
- \( \{ [r + s, 0, 2r + 3s]^T \mid r, s \in \mathbb{R} \} \) is a subspace of \( \mathbb{R}^3 \).

**THEOREM (spans are subspaces)**

The span of any (nonempty) set of vectors of \( \mathbb{R}^m \) is a subspace of \( \mathbb{R}^m \).

**Proof.** Call \( W \) the set \( \text{span}\{u_1, u_2, \ldots, u_k\} \). Then:
4. LINEAR SUBSPACES

- \( \mathbf{0} \in W \), because \( \mathbf{0} \) is always a linear combination of any nonempty set of vectors (just take all coefficients to be zero).
- Suppose some two vectors \( \mathbf{u}, \mathbf{v} \) belong to \( W \); we must show that \( \mathbf{u} + \mathbf{v} \) also does. Well, since these vectors are in \( W \), which is the span of the vectors \( \mathbf{u}_i \), they must be linear combinations of these vectors, that is, there must exist scalars \( a_i, b_j \) such that
  \[
  \mathbf{u} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \cdots + a_k \mathbf{u}_k \\
  \mathbf{v} = b_1 \mathbf{u}_1 + b_2 \mathbf{u}_2 + \cdots + b_k \mathbf{u}_k
  \]
  But then their sum is also a linear combination of them:
  \[
  \mathbf{u} + \mathbf{v} = (a_1 + b_1) \mathbf{u}_1 + (a_2 + b_2) \mathbf{u}_2 + \cdots + (a_k + b_k) \mathbf{u}_k
  \]
  proving that \( \mathbf{u} + \mathbf{v} \in W \).
- Suppose some \( \mathbf{u} \in W \), and let \( r \) be a scalar. We must prove that \( r \mathbf{u} \in W \). Similarly to the above, since \( \mathbf{u} \) is in \( W \), it must be a linear combination of the \( \mathbf{u}_i \)'s, and then \( r \mathbf{u} \) is equal to that same combination, just with all coefficients multiplied by \( r \). So \( r \mathbf{u} \) is also a combination of the \( \mathbf{u}_i \)'s, proving it is in \( W \).

- Use the result to show that
  \[
  \left\{ \begin{array}{c} a + d + e \\
  2b + d + 3e \\
  b - c - e \end{array} \right\} \mid a, b, c, d, e \in \mathbb{R}
  \]
  is a subspace of \( \mathbb{R}^3 \).
- Same for
  \[
  \left\{ \begin{array}{c} x \\
  y \\
  z \end{array} \right\} \mid y = 0, x = 5z
  \]
- Same for
  \[
  \left\{ \begin{array}{c} x \\
  y \\
  z \end{array} \right\} \mid x + y + z = 0
  \]

**DEFINITION (row space)**

The row space of a matrix \( A_{m \times n} \) is the subspace \( \text{Row}(A) \subseteq \mathbb{R}^n \) spanned by the vectors that form its rows. Note that each row has as many entries as the number of columns (\( n \)) of \( A \), so that’s why \( \text{Row}(A) \subseteq \mathbb{R}^n \).

- The row space of
  \[
  \begin{bmatrix}
  1 & 2 & 3 \\
  0 & 1 & 1 \\
  \end{bmatrix}
  \]
  is spanned by...
- The row space of
  \[
  \begin{bmatrix}
  1 & 2 \\
  0 & 0 \\
  -2 & -4 \\
  3 & 6 \\
  \end{bmatrix}
  \]
  is spanned by... (draw picture).

**THEOREM (Spaces associated to a matrix are subspaces)**

Let \( A_{m \times n} \) be a matrix. Then \( \text{Col}(A) \subseteq \mathbb{R}^m \), \( \text{Row}(A) \subseteq \mathbb{R}^n \) and \( \text{Null}(A) \subseteq \mathbb{R}^n \) are subspaces.

*Proof.* The row space and the column space are spans (of the sets of rows and columns for \( A \), respectively), so they are subspaces. The null space is the set of solutions to a homogeneous system, which we already know is always the span of some set of vectors (the ones that come up in the general solution, multiplying the free variables), so it’s also a subspace.

But an alternate, direct proof that \( \text{Null}(A) \) is a subspace can be helpful to learn how to verify the 3 properties of a subspace. \( \text{Null}(A) \) is defined as the set of all vectors \( \mathbf{u} \) such that \( A \mathbf{u} = \mathbf{0} \); let’s verify that this set satisfies the required properties:
- \( \mathbf{0} \in \text{Null}(A) \) because it satisfies the defining property of this set: \( A \mathbf{0} = \mathbf{0} \).
- Suppose \( \mathbf{u}, \mathbf{v} \in \text{Null}(A) \) (that is, \( A \mathbf{u} = \mathbf{0} \) and \( A \mathbf{v} = \mathbf{0} \)); we must use this information to show that \( \mathbf{u} + \mathbf{v} \in \text{Null}(A) \) (that is, \( A(\mathbf{u} + \mathbf{v}) = \mathbf{0} \)). But that is a simple consequence of the distributive property:
  \[
  A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}
  \]
• We leave this item as an exercise. You must assume that some $u \in \text{Null}(A)$, write down what that means, and then use that information to prove that, if $r$ is any scalar, then $ru \in \text{Null}(A)$.

**PROBLEM.** Consider the matrix

\[
A = \begin{bmatrix}
0 & 1 & 2 \\
3 & 1 & 0 \\
3 & 2 & 2
\end{bmatrix}
\]

Determine if each of the following is true or false:

a) $[3, 0, -2]^T \in \text{Row}(A)$
b) $[2, 2, 4]^T \in \text{Col}(A)$
c) $[-2, 6, -3]^T \in \text{Null}(A)$
d) $[1, 1, 1]^T \in \text{Row}(A)$
e) $[1, 1, 1]^T \in \text{Col}(A)$
f) $[1, 1, 1]^T \in \text{Null}(A)$

**Generating sets for certain subspaces**

We already mentioned that the span of a set of vectors is always a subspace. Those vectors can be called a **generating set for that subspace**.

Reciprocally, every subspace has a generating set for it, that is, every subspace is the span of some set of vectors. This is trivially true, because we can always take that set to be the entire subspace, an infinite set. But it’s always better to find only a few vectors that span the subspace. And we already know how to do it for the subspaces associated to a matrix $A$, since chapter 2:

- The subspaces $\text{Col}(A)$ and $\text{Row}(A)$ are spanned by the columns and rows of $A$, by definition.
- To find a generating set for $\text{Null}(A)$, use the vectors that show up in the vector form of the solution to the homogeneous system $Ax = 0$.

**PROBLEM.** Find generating sets for the row space, column space and null space:

\[
\begin{bmatrix}
1 & 2 & -1 \\
-1 & -3 & 4
\end{bmatrix}
\]

**PROBLEM.** Find generating sets for range and kernel:

\[
T[x, y]^T = [x - y, 2x - 2y, 3x - 3y]^T
\]

**Suggested problems from the book**

- Systematic: 1-42, 81-94
- True or false: 43-62
- Conceptual: 63-70, 75-76, 95-99

---

### 4.2. Basis and dimension

Key concepts: **basis, dimension, conditions for a set of vectors in a subspace to be linearly independent and/or a generating set for it**

Suppose we have a subspace $W$ and we found a generating set for it. We know that it can possibly be made smaller by removing vectors that are linear combinations of the other vectors. But once we reach a set of vectors that is independent, nothing else can be removed without altering the span (that is, without changing the fact that the vectors generate $W$). So there is something special about a linearly independent generating set for a subspace.
DEFINITION (basis)

Let \( W \subseteq \mathbb{R}^m \) be a subspace different from \( \{0\} \). A **basis** for \( W \) is a linearly independent generating set for \( W \), that is, a linearly independent set \( \{u_1, u_2, \ldots, u_k\} \) such that \( \text{span}\{u_1, u_2, \ldots, u_k\} = W \).

The plural of “basis” is written “bases”.

Subspaces other than \( \{0\} \) always have infinitely many bases.

- \( \{e_1, e_2, \ldots, e_m\} \subseteq \mathbb{R}^m \) is a basis for the trivial subspace \( \mathbb{R}^m \) of \( \mathbb{R}^m \).
- \( \{[1,2,0]^T, [3,1,0]^T\} \) is a basis for the subspace \( W \) of \( \mathbb{R}^3 \) consisting of the vectors with a zero third entry:
  \[ W = \{[x_1, x_2, 0]^T \mid x_1, x_2 \in \mathbb{R} \} \]

Another possible basis for \( W \) is \( \{[1,0,0]^T, [0,1,0]^T\} \).

- Find a basis for \( \text{Col}(A) \) where
  \[ A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 4 & 1 & 1 \\ 3 & 6 & 0 & 3 \\ 4 & 8 & 3 & 1 \end{bmatrix} \]

  (The pivot columns, because they generate the column space and are independent)

ALGORITHM (finding a basis if we have a generating set)

Suppose we have a generating set \( \{u_1, u_2, \ldots, u_k\} \) for some subspace \( W \):
\[ W = \text{span}\{u_1, u_2, \ldots, u_k\} \]

To find a basis for \( W \), take only the pivotal columns of the matrix
\[ A = \begin{bmatrix} u_1 & u_2 & \cdots & u_k \end{bmatrix} \]

(you’ll need to perform GE to find which columns are pivotal).

Justification. Due to the Column Correspondence Principle, each nonpivotal column can be eliminated from the generating set without altering the span, because they are combinations of the pivotal columns that come before them in \( A \). But the set containing all pivotal columns is independent.

PROBLEM. Find a basis for
\[ \text{span} \left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 4 \\ -6 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 6 \\ 3 \end{bmatrix} \right\} \]

PROBLEM. Find a basis for \( \text{Row}(A) \) where
\[ A = \begin{bmatrix} 5 & 3 & 1 \\ 3 & 1 & 4 \\ 2 & 2 & -3 \end{bmatrix} \]

We already know that, for a subset \( S \subseteq \mathbb{R}^m \):
- If \( S \) is a generating set for \( \mathbb{R}^m \), then \( S \) must contain at least \( m \) vectors.
- If \( S \) is independent, then \( S \) must contain at most \( m \) vectors.

Therefore, any basis of \( \mathbb{R}^m \) contains exactly \( m \) vectors. It turns out an analogous result is true for any subspace, and it gives mathematical meaning to something we all know intuitively: a line is 1-dimensional, a plane is 2-dimensional, the space is 3-dimensional.
4.2. BASIS AND DIMENSION

**THEOREM (definition of dimension)**

See Theorem 4.5 on page 245 for a proof.

Let \( W \subseteq \mathbb{R}^m \) be a subspace, \( W \neq \{0\} \). Then all bases for \( W \) contain the same number of vectors. That number is called the **dimension** of \( W \), denoted \( \text{dim}(W) \). We also define \( \text{dim}(\{0\}) \) to be 0.

In particular, \( \text{dim}(\mathbb{R}^m) = m \).

**PROBLEM.** Find the dimension of the following subspace of \( \mathbb{R}^4 \):

\[
W = \{ [x_1, x_2, x_3, x_4]^T \mid x_2 - 3x_3 = x_1 + x_2 - x_4 = 0 \}
\]

**PROBLEM.** Find the dimensions of the row space, column space and null space of the following matrices:

\[
A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 2 & -2 \end{bmatrix}
\]

**THEOREM (reduction and extension)**

See also Theorems 4.3 and 4.4 on pages 243-245. In the book they come before the previous theorem and are used to prove it.

Let \( W \subseteq \mathbb{R}^m \) be a subspace.

a) Given any generating set for \( W \), there is a basis for \( W \) **contained** in it.

b) Given any set of linearly independent vectors in \( W \), there is a basis for \( W \) **containing** it.

Therefore, a basis for a subspace can be thought of as a generating set with the fewest possible vectors, or as a linearly independent set with the most possible vectors.

This theorem justifies the following algorithm, which is useful when we want to know whether a given set \( S \) is a basis for a given subspace \( W \). This is important because it’s not like we could have simply tried to find a basis for \( W \) and to see whether it’s \( S \) or not (\( W \) has infinitely many possible bases).

**ALGORITHM (confirming that a set is a basis for a given subspace)**

(look at the box in the end of page 248 and beginning of 249 in the book)

To show that a set \( S \) is a basis for a subspace \( W \), verify 3 things:

- That \( S \subseteq W \).
- That \( S \) is linearly independent.
- That \( S \) contains the correct number of vectors, that is, \( \text{dim}(W) \) vectors.

**Justification.** Once we verified the first two items, because of the previous theorem (the extension part), there must exist a basis for \( W \) containing \( S \) (it could be \( S \) itself or some larger set). But then the third item tells us that \( S \) already has the required number of vectors, so nothing strictly larger than \( S \) can possibly be a basis. \( S \) itself must be the extension that is a basis.

- Show that \( \{ [1, -1, 1, 0]^T, [1, 0, 1, -1]^T, [0, 1, 1, -1]^T \} \) is a basis for \( \{ [x_1, x_2, x_3, x_4]^T \mid x_1 + x_2 + x_4 = 0 \} \).
- Show that \( \{ [1, -2, 0]^T, [1, 0, -3]^T \} \) is a basis for \( \text{Null} \begin{bmatrix} 2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \).

**The size of a set can determine linear dependence or non-generation**

Let \( W \subseteq \mathbb{R}^m \) be a subspace, let \( S \subseteq W \) be a set of vectors. As a simple consequence of the reduction and extension theorems:

- If \( S \) has fewer than \( \text{dim}(W) \) vectors, it cannot be a generating set for \( W \).
- If \( S \) has more than \( \text{dim}(W) \) vectors, it cannot be linearly independent.

- \( W = \text{span} \{ [1, 1, 1]^T, [1, 2, 0]^T, [0, 3, 0]^T \} \) is not spanned by \( \{ [1, 1, 1]^T, [2, 3, 1]^T \} \).
• The set \([1, 2, 3]^T, [3, 2, 1]^T, [5, 6, 7]^T\), contained in the subspace \(\{x_1, x_2, x_3|^T \mid 2x_2 = x_1 + x_3\}\), must be linearly dependent.

Suggested problems from the book
- Systematic: 1-32, 59-66
- True or false: 33-52
- Conceptual: 53-58, 67-70

4.3. Dimension of subspaces associated to a matrix

Key concepts: finding bases and calculating the dimension of subspaces associated to matrices and linear transformations

Dimension of column space and null space

Let \(A\) be a matrix. Then
- \(\dim(\text{Col}(A)) = \text{rank}(A)\)
- \(\dim(\text{Null}(A)) = \text{nullity}(A)\)

*Justification.* Since a basis for the column space is given by the pivotal columns, the dimension of this subspace is the number of such columns, which is the rank. On the other hand, \(\text{Null}(A)\) is the solution set to the homogeneous system \(Ax = 0\), which we know can be written as the span of the vectors that appear in the vector form solution; we also already remarked that these vectors are always independent, so they are a basis, and the number of such vectors is the number of free variables, which is the nullity.

This allows us to compute the dimension of a subspace if we realize that this subspace is a null space:
- Find the dimension of \(\{[x_1, x_2, \ldots, x_5]^T \mid x_1 - 2x_4 + x_5 = x_2 + x_3 + x_5 = 0\}\).
- Find the dimension of \(\{[x_1, x_2, x_3]^T \mid 3x_1 - 2x_3\}\).

To find a basis for the row space of a matrix, we can always consider the transposed matrix and find a basis for its column space. But there’s an alternative method:

**THEOREM (rows of RREF are basis for row space)**

The nonzero rows of the RREF form of a matrix \(A\) are a basis for \(\text{Row}(A)\).

*Proof sketch.* Elementary row operations do not change the row space of a matrix. Indeed, the way that these operations change the vectors is by replacing them with linear combinations of the others, so that the set generated by them never gains anything new. It doesn’t lose anything either because the operations are reversible.

Then the subspaces \(\text{Row}(A)\) and \(\text{Row}(R)\), where \(R\) is the RREF of \(A\), are the same. But \(\text{Row}(R)\) is spanned by the rows of \(R\), and we may eliminate the ones that are identically zero. The ones that are left are independent, as we remarked in chapter 1, so we have a basis.

**Find a basis for \(\text{Row}(A)\) where**

\[
A = \begin{bmatrix}
3 & 1 & -2 & 1 & 5 \\
1 & 0 & 1 & 0 & 1 \\
-5 & -2 & 5 & -5 & -3 \\
-2 & -1 & 3 & 2 & -10
\end{bmatrix}
\]
If $R$ given below is the RREF form of a matrix $A$, determine the dimensions of $\text{Col}(A)$, $\text{Null}(A)$, $\text{Row}(A)$ and $\text{Null}(A^T)$. Can we find what each of these subspaces are?

$$R = \begin{bmatrix}
1 & 0 & 4 & 0 \\
0 & 1 & 4 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}$$

We finish with two theorems that are frequently used to prove other theorems (but not so much in calculations or algorithms).

**THEOREM**

Let $A$ be any matrix, let $T : \mathbb{R}^n \to \mathbb{R}^m$ be any linear transformation. Then:

a) Column space and row space always have the same dimension: $\dim(\text{Col}(A)) = \dim(\text{Row}(A))$.

b) Matrix and its transpose always have the same rank: $\text{rank}(A) = \text{rank}(A^T)$.

c) The rank-nullity theorem: $\dim(\text{Range}(T)) + \dim(\text{Ker}(T)) = \dim(\text{Domain}(T))$.

**Proof.**

a) Since a basis for $\text{Row}(A)$ is given by the nonzero rows of its RREF form $R$, we have that $\dim(\text{Row}(A))$ is the number of rows with a pivot. But that’s the same as the number of columns with a pivot, since each pivot belongs to a different row and a different column. And that number is $\dim(\text{Col}(A))$, since $\text{Col}(A)$ has a basis given by the pivotal columns of $A$.

b) The rank is the dimension of the column space, but we have just shown that this is the same as the dimension of the row space. Since the transpose turns columns into rows, we have $\text{Col}(A) = \text{Row}(A^T)$.

In particular the dimensions of these spaces, $\text{rank}(A)$ and $\text{rank}(A^T)$, are the same.

c) Let $A$ be the standard matrix of $T$. The range is the column space, whose dimension is the number of pivotal columns of $A$. The kernel is the null space, whose dimension is the number of nonpivotal columns of $A$. Their sum is the total number of columns of $A$, which is the $n$ that appears in the domain $\mathbb{R}^n$ of $T$, that is, the dimension of the domain.

**THEOREM (subspace of a subspace)**

Let $V, W \subseteq \mathbb{R}^m$ be two subspaces. Then:

a) If $V \subseteq W$, then $\dim(V) \leq \dim(W)$.

b) If $V \subseteq W$ and $\dim(V) = \dim(W)$, then $V = W$.

**Proof.**

a) Choose a basis $B$ for $V$. It is formed by linearly independent vectors all contained in $W$ by assumption, so it can be extended to a basis for $W$, which will then have at least as many vectors as $B$.

b) Choose a basis $B$ for $V$. It is formed by linearly independent vectors all contained in $W$ by assumption, and it also has the correct number of elements to be a basis for $W$, also by assumption. Then $B$ spans both $V$ and $W$, hence $V = W$.

Look at both tables on page 259 in the book (summary on dimensions of subspaces associated to a matrix and how to find bases for them)

**Suggested problems from the book**

- Systematic: 1-40
- True or false: 41-60
- Conceptual: 72, 74
4.4. Coordinate systems

Key concepts: coordinates of a vector in a basis, conic sections, rotations of the standard coordinates of the plane

**Motivation for coordinate changes**

Suppose we want to draw the curve on the plane with equation

\[5x^2 + 2xy + 5y^2 = 12\]

Somehow we realize that it becomes simpler when written in terms of new variables

\[
\begin{align*}
  u &= x + y \\
  v &= x - y
\end{align*}
\]

Indeed, first solve for the old variables \((x, y)\) in terms of the new variables \((u, v)\):

\[
\begin{align*}
  x &= \frac{1}{2}(u + v) \\
  y &= \frac{1}{2}(u - v)
\end{align*}
\]

then plug these values into the equation for the curve and simplify:

\[
5\left(\frac{u + v}{2}\right)^2 + 2\left(\frac{u + v}{2}\right)\left(\frac{u - v}{2}\right) + 5\left(\frac{u - v}{2}\right)^2 = 12
\]

\[
5u^2 + 10uv + 5v^2 + 2u^2 - 2v^2 + 5u^2 - 10uv + 5v^2 = 48
\]

\[
12u^2 + 8v^2 = 48 \\
3u^2 + 2v^2 = 12
\]

What object does this equation describe? The equation \(3x^2 + 2y^2 = 12\) would have been an ellipse, symmetrical with respect to the \(x, y\) axes, that is more tall than it is wide (see below for details).

Picture!

So the equation that we actually have, \(3u^2 + 2v^2 = 12\), can be drawn according to the same description once we know where to place the \(u, v\) axes. The \(u\) axis is the line with equation \(v = 0\) (the points along the \(u\)-axis have value 0 for their \(v\) coordinate), that is, \(x - y = 0\) or \(y = x\). Similarly the \(v\) axis is the line \(y = -x\). Then we see that the \((u, v)\) coordinate system is just a rotation (and rescaling) of the usual \((x, y)\) system.

Picture!

So here is the curve that we wanted to draw:

Picture!

---

**DEFINITION (coordinates of a vector with respect to a basis)**

Let \(W \subseteq \mathbb{R}^m\) be a subspace, let \(B = \{u_1, u_2, \ldots, u_k\} \subseteq W\) be a basis for \(W\). Then any vector \(v \in W\) can be expressed in a unique way as a linear combination of the vectors in \(B\):

\[v = c_1 u_1 + c_2 u_2 + \ldots + c_k u_k\]
The scalars $c_i$ receive the name of the coordinates of $v$ in the basis $B$, and we use a special symbol for the vector containing them, called the coordinate vector of $v$:

$$[v]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$$

**Justification.** Any vector $v \in W$ can be expressed as a linear combination of the basis vectors in at least one way, because $B$ spans $W$. But there can be only one way to do it, because those vectors are also independent (that is, the system $[u_1, u_2, \ldots, u_k]x = v$ has no free variables).

- Let $B$ be the basis $\{[1, 2, -3]^T, [0, 1, -1]^T\}$ of the subspace $W = \{[x_1, x_2, x_3]^T \mid x_1 + x_2 + x_3 = 0\}$. Then:
  - For $u = [0, 2, -2]^T$, the coordinates are $[u]_B = [0, 2]^T$.
  - For $v = [-1, -3, 4]^T$, the coordinates are $[v]_B = [-1, -1]^T$.
  - The coordinates of $\mathbf{0}$ are $[\mathbf{0}]_B = [0, 0]^T$.
  - The vector $w = [1, 2, 3]^T$ doesn’t have any coordinates with respect to $B$ because it is not in $W$.
  - If $x$ is the vector such that $[x]_B = [\pi, 3\pi]^T$, then it is
    $$x = \pi \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} + 3\pi \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \pi \\ 5\pi \\ -6\pi \end{bmatrix}$$

- The coordinates of a vector in $\mathbb{R}^m$ with respect to the standard basis $\{e_1, e_2, \ldots, e_m\}$ are just the entries of the vector:
  
  $$\begin{bmatrix} 4 \\ \pi^3 \\ 0 \\ 9 \end{bmatrix} = 4e_1 + \pi^3e_2 + 0e_3 + 9e_4 = 4 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \pi^3 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + 9 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

That is, if $E$ denotes the standard basis, then

$$[v]_E = v \quad \text{for all } v$$

- Draw the basis $B = \{[1, 3]^T, [-1, 1]^T\}$ of $\mathbb{R}^2$, then compute and draw the vectors whose coordinates with respect to $B$ are $[1, 1]^T$, $[3, 0]^T$, $[-2, 1]^T$.

**PROBLEM.** Find $[u]_B$ if $u = [1, 2]^T$ and $B$ is the basis $\{[1, 1]^T, [1, -1]^T\}$ of $\mathbb{R}^2$.

**PROBLEM.** Find $v$ if $[v]_B = [5, 3]^T$ and $B$ is the basis $\{[1, 0, 2]^T, [2, 0, 1]^T\}$ of a subspace of $\mathbb{R}^3$.

**PROBLEM.** Suppose that a subspace $W \subseteq \mathbb{R}^6$ has a basis $B = \{u_1, u_2, u_3, u_4\}$. The set $C = \{u_3, 2u_1, u_4, -5u_2\}$ is also a basis for $W$. If $v \in W$ is such that $[v]_B = [2, 7, 3, 0]^T$, then what is $[v]_C$?

We see from the problems above that it’s always easy to find $v$ when we know $[v]_B$, but to solve a problem that gives $v$ and asks for $[v]_B$ we need to solve a linear system. There’s also an alternative formula for solving such problems (but in most cases the amount of work is the same):

**ALGORITHM (finding the coordinate vector in a given basis)**

Let $B = \{u_1, u_2, \ldots, u_m\}$ be a basis for $\mathbb{R}^m$. To find the coordinates of a vector $v \in \mathbb{R}^m$ in the basis $B$:

- Construct the matrix $B$ that has the basis vectors as columns:
  $$B = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$$

- Find $B^{-1}$ (the matrix $B$ will always be invertible, since its columns are independent).
The coordinate vector is given by:

\[ [v]_B = B^{-1}v \]

**Justification:** It’s enough to check that this works when \( v \) is one of the vectors in the basis \( B \) (because they span the whole space and the equation is linear in \( v \), one can then prove that it must be true for all vectors).

So let’s justify it for the vectors \( v_i \). The coordinates of a particular \( v_i \) in the basis \( B \) itself are easy to find: they are all 0, except for a 1 in the slot corresponding to \( v_i \) itself:

\[
v_i = 0v_1 + 0v_2 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + \cdots + 0v_m
\]

That is, the vector containing these coordinates is the \( i \)-th standard vector:

\[ [v_i]_B = e_i \]

Do we obtain this by applying the formula claimed in the algorithm? That is, is it true that

\[ B^{-1}v_i = e_i \]

This is indeed true. It follows by calculating each column separately in the following equation:

\[ B^{-1}B = I \]

(remember, the columns of \( I \) are the standard vectors \( e_i \), and the columns of the matrix product \( B^{-1}B \) are obtained by the matrix-vector product \( B^{-1}b_i \)).

**PROBLEM.** Let \( B = \{[1, 1, 0]^T, [1, 1, 1]^T, [3, 2, 1]^T\} \) be a basis for \( \mathbb{R}^3 \). Find the coordinates of the vector \( u = [1, 2, 3]^T \) in \( B \) by the above formula.

---

**DEFINITION (standard ellipses)**

**Standard ellipse**: The set of solutions \( (x, y) \) to an equation of the form

\[
\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1
\]

where \( a, b > 0 \). This is a stretched circle (or a perfect circle when \( a = b \)) centered at the origin:

\[ x^2 + y^2 = 1 \]

Draw the ellipse \( 4x^2 + 3y^2 = 8 \).

**DEFINITION (coordinate systems in the plane)**

_This definition is not found in the book, at least in this form._

Let \( (x,y) \) denote the usual coordinates of the plane. We call \( (u,v) \) defined by equations of the form

\[
\begin{align*}
\begin{cases}
u &= ax + by \\
v &= cx + dy
\end{cases}
\end{align*}
\]

a **new coordinate system** if the matrix

\[
U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}
\]

The reason for studying how a vector is expressible in terms of a basis different from the standard basis is that the description of some equation or geometric object in one basis may be simpler than in the standard basis. For now, let’s learn how to get the complicated equation of a rotated ellipse from the simple equation of the unrotated ellipse. In chapter 6 we’ll see how to do the reverse: given a complicated equation of a rotated ellipse, find the rotated coordinate system that simplifies it.

_The book also considers parabolas and hyperbolas in this chapter, but I feel like ellipses are the easiest of the 3 to understand and already give enough practice with rotating coordinate systems._

---

**Picture!**
of the coefficients is invertible (so that every \((x, y)\) point corresponds to a value of \((u, v)\) and vice-versa).

The associated linear transformation

\[
T_U : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \; ; \; T_U \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}
\]

then tells us how to obtain the coordinates of a point in the new system in terms of its usual coordinates:

\[
\begin{bmatrix} u \\ v \end{bmatrix} = T_U \begin{bmatrix} x \\ y \end{bmatrix}
\]

Because \(U\) is invertible, this transformation is invertible and we can also find the usual coordinates in terms of the new:

\[
\begin{bmatrix} x \\ y \end{bmatrix} = T_U^{-1} \begin{bmatrix} u \\ v \end{bmatrix}
\]

- Show that the formulas
  \[
  \begin{align*}
  u &= 2x - y \\
  v &= x + 3y
  \end{align*}
  \]
  define a new coordinate system in the plane. Rewrite the equation of the line \(y = -x + 4\) in the new coordinates. Rewrite the equation of the line \(v = 2u + 5\) in the old coordinates.

- What is the equation of the ellipse \(2x^2 + y^2 = 1\) in a coordinate system rotated 30° counterclockwise?

- Draw the ellipse \(\frac{9}{2} x^2 + \sqrt{3}xy + \frac{7}{2} y^2 = 1\) by considering a new coordinate system rotated 60° clockwise.

**Remark (more general coordinate systems)**

There are many examples of situations, not only in the plane but also in the general \(\mathbb{R}^m\), where a new system of coordinates simplifies some calculation, but it’s not necessarily a rotated system or even a system defined by a linear transformation. The goal of this section was solely to introduce the topic of change-of-coordinates in the simplest possible context.

**Suggested problems from the book**

- **Systematic:** 1-30, 55-58, 67-70, 79, 82, 84, 86, 87-94
- **True or false:** 31-47
- **Conceptual:** 51-54, 59-66, 71-78, 95-98, 102

### 4.5. Matrix representations of linear transformations in other bases

**Key concepts:** matrix of a linear transformation in other bases, linear operators, similar matrices

**Motivation**

The linear transformation \(T : \mathbb{R}^2 \rightarrow \mathbb{R}^2\) defined by the matrix

\[
\begin{bmatrix}
-1 & 0 \\
0 & 1
\end{bmatrix}
\]

has a simple geometric interpretation: it is the reflection around the \(y\)-axis (because it transforms a vector \([x_1, x_2]^T\) into \([-x_1, x_2]^T\), and flipping the \(x\)-coordinate means bringing the vector across the \(y\)-axis).

Now consider the transformation \(U : \mathbb{R}^2 \rightarrow \mathbb{R}^2\) defined by the matrix

\[
A = \begin{bmatrix}
-3/5 & -4/5 \\
-4/5 & 3/5
\end{bmatrix}
\]
At first it may look more complicated than \( T \) above. But consider what it does to the vectors \( u_1 = [2, 1]^T \) and \( u_2 = [-1, 2]^T \) (which form a basis of \( \mathbb{R}^2 \) and are perpendicular to each other just like the standard vectors \( e_1, e_2 \)):

Picture!

We can calculate and see that \( U(u_1) = [-2, -1]^T = -u_1 \) and \( U(u_2) = [-1, 2]^T = u_2 \). Then, by linearity, a general vector \( v \in \mathbb{R}^2 \) written in the basis \( B = \{ u_1, u_2 \} \) as

\[ v = c_1 u_1 + c_2 u_2 \]

gets transformed into

\[ U(v) = c_1 U(u_1) + c_2 U(u_2) = -c_1 u_1 + c_2 u_2 \]

That is, it first coordinate in the basis \( B \) gets flipped, just like happened with \( T \) above. So we see that \( U \) is the reflection around the axis defined by the vector \( u_2 \).

This shows that, if we want to encode the transformation \( U \) into a matrix, it makes more sense to use not its standard matrix \( A \), but the matrix that tells us how the coefficients of a general vector in the basis \( B \) get transformed, which would be

\[
\begin{bmatrix}
-1 & 0 \\
0 & 1
\end{bmatrix}
\]

This way, instead of the more complicated definition

\[
U(v) = \begin{bmatrix}
-3/5 & -4/5 \\
-4/5 & 3/5
\end{bmatrix} v
\]

we have the simpler

\[
[U(v)]_B = \begin{bmatrix}
-1 & 0 \\
0 & 1
\end{bmatrix} [v]_B
\]

DEFINITION (matrix of a linear transformation with respect to given bases)

Let \( T : \mathbb{R}^n \to \mathbb{R}^m \) be a linear transformation. Suppose \( B = \{ b_1, b_2, \ldots, b_n \} \subseteq \mathbb{R}^n \) is a basis for its domain and \( C = \{ c_1, c_2, \ldots, c_m \} \subseteq \mathbb{R}^m \) is a basis for its codomain. The **matrix of \( T \) with respect to the bases \( B \) and \( C \)**, denoted \([T]_{B,C}\), is the matrix whose columns contain the vectors \( T(b_1), T(b_2), \ldots, T(b_n) \) expressed in the basis \( C \):

\[
[T]_{B,C} = \begin{bmatrix}
[T(b_1)]_C & [T(b_2)]_C & \cdots & [T(b_n)]_C
\end{bmatrix}
\]

Note that this is an \( m \times n \) matrix, just like the standard matrix for \( T \).

Just like the standard matrix \( A \) of a linear transformation \( T \) is the matrix that defines \( T \) in the sense of

\[
T(u) = Au
\]

the matrix \([T]_{B,C}\) of \( T \) with respect to some bases \( B \) and \( C \) is the matrix that defines it on vectors written in those bases:

\[
[T(u)]_C = [T]_{B,C}[u]_B
\]

In the textbook, this topic is discussed only in the case \( m = n \) and with \( B = C \). In that case, instead of \([T]_{B,B}\), we can denote the matrix simply by \([T]_B\). But there’s no reason to restrict the topic to just this case.

- If \( T \) has standard matrix

\[
\begin{bmatrix}
1 & 2 & 0 \\
0 & 3 & 4
\end{bmatrix}
\]

and we consider the following bases for its domain and codomain:

\[
B = \left\{ \begin{bmatrix} 2 \\ 7 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \right\} \quad C = \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \end{bmatrix} \right\}
\]

then . . . Compute \( T[0, -1, 6]^T \) directly and by the formula as well.
• If $T$ has standard matrix $\begin{bmatrix} -4 & 3 \\ -6 & 5 \end{bmatrix}$, and we consider the following basis for its domain and codomain:

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

then …. Compute $T[4,5]^T$ directly and by the formula as well.

• Suppose $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, and let $B = \{u_1, u_2, u_3\}$ be a basis of $\mathbb{R}^3$. Suppose that

$$T(u_1) = u_1 + 3u_3 \quad T(u_2) = 0 \quad T(u_3) = u_1 + u_2 + u_3$$

What is $[T]_B$?

• Suppose $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ satisfies

$$T[0,2]^T = [2,2,0]^T \quad T[2,3]^T = [0,-5,-5]^T$$

If $B = \{[0,0]^T, [2,3]^T\}$ and $C = \{[1,1,0]^T, [0,1,0]^T, [0,1,1]^T\}$, then what is $[T]_{B,C}$?

The following remark, which was hinted at at the previous examples, gives a way to compute the matrix $[T]_{B,C}$ as the product of 3 matrices:

**ALGORITHM** (relation between standard matrix and matrix with respect to different bases)

Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear transformation, let $B \subseteq \mathbb{R}^n$ and $C \subseteq \mathbb{R}^m$ be bases for its domain and codomain. We have previously studied some matrices associated to these objects:

- Associated to $T$ there is the standard matrix $A_{m \times n} = \begin{bmatrix} T(e_1) & T(e_2) & \cdots & T(e_n) \end{bmatrix}$.
- Associated to $B$ there is the matrix $B_{n \times n} = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}$.
- Associated to $C$ there is the matrix $C_{m \times m} = \begin{bmatrix} c_1 & c_2 & \cdots & c_m \end{bmatrix}$.

It turns out that the matrix of $T$ in the bases $B$ and $C$ is expressible in terms of them:

$$[T]_{B,C} = C^{-1}AB$$

*Justification.* Column $i$ of $[T]_{B,C}$ is given by $T(b_i)$ written in basis $C$, which we know (from the previous section) can be computed using the matrix $C$ as $C^{-1}T(b_i)$. We also know that $T(b_i)$ is computed using the matrix $A$ as $Ab_i$ (according to the definition of what the standard matrix is), which is the $i$-th column of $AB$.

Apply this to the first two examples given above.

**DEFINITION** (linear operator)

The word **operator** (or **linear operator**) is used to designate a linear transformation whose codomain and domain coincide:

$$T : \mathbb{R}^m \rightarrow \mathbb{R}^m$$

Its standard matrix $A$ is then square, as is the matrix $[T]_{B,B}$ with respect to any two given bases $B, C$ of $\mathbb{R}^m$.

For operators, it makes sense to consider the matrix $[T]_{B,B}$, with respect to one and the same basis (in the textbook, the notation for it is simply $[T]_B$).

**PROBLEM.** Consider the operator $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by $Tu = 2u$. What is its matrix with respect to the standard basis of $\mathbb{R}^2$ rotated by $45^\circ$ counterclockwise?

**PROBLEM.** Consider the following basis of $\mathbb{R}^2$:

$$B = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right\}$$

The linear operator defined by

$$\left[ T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right]_B = \left[ \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \right]_B$$
has the geometric interpretation of a projection onto the line spanned by \([1, -1]^T\) (because it kills off the component of the vector with respect to the vector \([-1, -1]^T\)). Find its standard matrix as well as \([T]_B\).

**DEFINITION (similar matrices)**

Two square matrices \(A, B\) of the same dimensions are said to be similar if there exists an invertible matrix \(P\) such that

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

Or equivalently, if there exists an invertible matrix \(Q\) such that

\[
B = QAQ^{-1}
\]

(simply take \(P = Q^{-1}\) to see that the two equations are equivalent). In this section it’s preferrable to use the first formulation; in the next chapter it will be more convenient to use the second.

We will learn in the next chapter a good way to decide if two matrices are similar or not. For now, let’s just be content with knowing where this concept comes from:

**THEOREM (matrices of an operator in different bases are similar)**

Let \(T : \mathbb{R}^m \rightarrow \mathbb{R}^m\) be an operator.

a) Given any basis \(B \subseteq \mathbb{R}^n\), the standard matrix \(A\) of \(T\) and the matrix \([T]_B\) are similar.

b) More generally, if \(C \subseteq \mathbb{R}^m\) is some other basis of \(\mathbb{R}^m\), then \([T]_B\) and \([T]_C\) are also similar.

**Proof.**

a) We just learned that there is a relation between \(A\) and \([T]_B\) involving the matrix \(B\) of the basis \(B\):

\[
[T]_B = B^{-1}AB
\]

Then \(A\) and \([T]_B\) are similar, with \(B\) playing the role of the invertible matrix \(P\) from the definition.

b) Similarly, we have (for \(C\) being the matrix of the basis \(C\))

\[
[T]_C = C^{-1}AC
\]

We can solve for \(A\) in here by multiplying both sides by \(C\) on the left and \(C^{-1}\) on the right:

\[
C[T]_C C^{-1} = A
\]

Then plug this into the equation for \([T]_B\) from item (a):

\[
[T]_B = B^{-1}(C[T]_C C^{-1})B = (B^{-1}C)[T]_C (C^{-1}B)
\]

This doesn’t look like the kind of relation that we want, which is

\[
[T]_B = P^{-1}[T]_C P
\]

for some invertible matrix \(P\). But it is in fact in this form, for \(P = C^{-1}B\); we just need to remember that taking the inverse of a product reverses its order:

\[
P^{-1} = B^{-1}C
\]

Thus \([T]_B\) and \([T]_C\) are similar.

---

Suggested problems from the book

- **Systematic:** 1-18, 47-54, 101
- **True or false:** 19-38
- **Conceptual:** 39-46, 63-64, 89, 103
Eigen means auto or self in German. Square matrices have certain special numbers associated to them, called their eigenvalues (e-values for short), and each e-value is associated to a subspace called the eigenspace (e-space for short). The vectors in that e-space are called eigenvectors (e-vectors for short). The definition is that, if we think of a square matrix as a linear operator, then an e-vector is a vector that is transformed in the simplest of ways: it just becomes a multiple of itself, and the e-value dictates which multiple. If we are able to find enough e-vectors for a given matrix, we can perform a change of variables that brings the matrix to a much simpler form to study. This process is called diagonalization, but it may not always be possible to carry out.

(We use the abbreviations “e-values”, “e-vectors”, “e-spaces”, but nobody actually calls them that when speaking. You should still say “eigenvalues”, etc.)

This entire chapter concerns only square matrices.

5.1. Eigenvalues and eigenvectors

Key concepts: eigenvalues, eigenvectors, eigenspaces, finding eigenvectors given the eigenvalue, showing that a scalar is or isn’t an eigenvalue

A square matrix can be thought of as a linear operator, that is, a linear transformation that takes vectors into the same Euclidean space where they came from:

\[ u \in \mathbb{R}^m \mapsto A u \in \mathbb{R}^m \]

So it makes sense to ask whether the vector got transformed into a multiple of itself.

**DEFINITION (eigenvectors and eigenvalues)**

Let \( A_{m \times m} \) be a square matrix. A nonzero vector \( u \in \mathbb{R}^m \) is called an eigenvector of \( A \) if there exists some scalar \( \lambda \) such that

\[ Au = \lambda u \]

In that case, the scalar \( \lambda \) is called the eigenvalue of \( A \) corresponding to \( u \).

This says that \( A \) transforms \( u \) into a multiple of itself. The letter \( \lambda \) is called “lambda” (the Greek equivalent of the letter L) and is the traditional symbol for an e-value.

\( 0 \) is not considered an e-vector because it would correspond to all real numbers as e-values: \( A 0 = \lambda 0 \) is true for any \( \lambda \). However, 0 as an e-value is allowed and does sometimes happen!

- Consider \( A = \begin{bmatrix} -2 & -4 & 2 \\ -2 & 1 & 2 \\ 4 & 2 & 5 \end{bmatrix} \). Then:
  - The vector \( [2, -3, -1]^T \) is an e-vector corresponding to e-value \( \lambda = 3 \).
  - The vector \( [4, -6, -2]^T \) is another e-vector corresponding to e-value \( \lambda = 3 \).
  - The vector \( [1, 6, 16]^T \) is an e-vector corresponding to e-value \( \lambda = 6 \).
  - The vector \( u = [1, 2, 3]^T \) is not an e-vector because \( Au \) is not a multiple of \( u \).

- Consider \( B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 7 \end{bmatrix} \). Then:
  - The vectors \( [0, 1]^T, [0, -1]^T, [0, \pi]^T \) are examples of e-vectors corresponding to e-value \( \lambda = 7 \).
  - The vectors \( [1, 0]^T, [5, 0]^T, [-1000, 0]^T \) are examples of e-vectors corresponding to e-value \( \lambda = 0 \).

- You are told that the vectors \( u = [2, 5]^T \) and \( v = [-1, -6]^T \) are e-vectors of the matrix \( C = \begin{bmatrix} 17 & -6 \\ 45 & -16 \end{bmatrix} \).

Find the corresponding e-values.
Eigenstuff of an operator

By definition, e-values and e-vectors of an operator are the same as those of its standard matrix. These concepts are not defined for a general linear transformation; it must be a linear operator because the corresponding matrix must be square.

- According to one of the examples given above, the operator
  \[
  T = \begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3
  \end{bmatrix} = \begin{bmatrix}
  -2x_1 - 4x_2 + 2x_3 \\
  -2x_1 + x_2 + 2x_3 \\
  4x_1 + 2x_2 + 5x_3
  \end{bmatrix}
  \]
  has an e-value \( \lambda = 3 \) with a possible e-vector \([2, -3, -1]^T\).

Given a square matrix \( A \), it turns out that very few scalars \( \lambda \) are e-values of \( A \). For most scalars \( \lambda \), there simply is no vector \( u \) (other than zero) satisfying \( Au = \lambda u \). The next section will be about determining which scalars are e-values. For now, we’ll only worry about finding e-vectors assuming we already know that \( \lambda \) is an e-value.

**THEOREM (definition of eigenspace)**

Let \( A_{m \times m} \) be a square matrix. Suppose \( \lambda \) is an e-value. The set
\[
E_\lambda = \{ u \in \mathbb{R}^m \mid Au = \lambda u \}
\]
of all corresponding e-vectors (plus the zero vector) is a subspace, called the eigenspace corresponding to \( \lambda \).

We call geometric multiplicity of \( \lambda \) the dimension \( \dim(E_\lambda) \).

**Proof.** This will follow from the next algorithm, which shows that \( E_\lambda \) is the null space of a matrix, and therefore a subspace. But it’s always helpful to see how to prove something is a subspace directly, by verifying the 3 properties:
- \( 0 \in E_\lambda \) because \( 0 \) satisfies the defining property of \( E_\lambda \):
  \[
  A0 = \lambda 0
  \]
  (Notice that we have explicitly said that \( 0 \) is never considered an e-vector, but \( E_\lambda \) is not just the set of all e-vectors corresponding to \( \lambda \); it is the set of those together with \( 0 \)).
- Suppose \( u, v \) satisfy the property defining \( E_\lambda \):
  \[
  Au = \lambda u \quad Av = \lambda v
  \]
  We must show that \( u + v \) also does. Try it!
- Suppose \( u \) satisfies the property above, and let \( r \) be any scalar. Then \( cu \) also does:
  \[
  A(cu) = c(Au) = c(\lambda u) = \lambda (cu)
  \]
  proving that \( cu \in E_\lambda \).

**ALGORITHM (finding e-vectors of an e-value \( \lambda \))**

Let
\[
A = \begin{bmatrix}
  a_{11} & a_{12} & a_{13} & \cdots & a_{1m} \\
  a_{21} & a_{22} & a_{23} & \cdots & a_{2m} \\
  a_{31} & a_{32} & a_{33} & \cdots & a_{3m} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mm}
  \end{bmatrix}
\]
be a square matrix. If \( \lambda \) is an e-value, then the corresponding e-vectors (and the vector \( 0 \)) are the solutions to the homogeneous system
\[
A = \begin{bmatrix}
  a_{11} - \lambda & a_{12} & a_{13} & \cdots & a_{1m} \\
  a_{21} & a_{22} - \lambda & a_{23} & \cdots & a_{2m} \\
  a_{31} & a_{32} & a_{33} - \lambda & \cdots & a_{3m} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mm} - \lambda
  \end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  \vdots \\
  x_m
  \end{bmatrix} = \begin{bmatrix}
  0 \\
  0 \\
  0 \\
  \vdots \\
  0
  \end{bmatrix}
\]
This can also be written as

$$E_\lambda = \text{Null}(A - \lambda I)$$

**Justification.** E-vectors (and 0) are the vectors $u$ satisfying

$$Au = \lambda u$$

If we write $\lambda u$ as $\lambda I u$, move it to the left side and factor the $u$, we get the equivalent equation

$$(A - \lambda I)u = 0$$

This is the homogeneous system above (the matrix $A - \lambda I$ is simply $A$ with a $\lambda$ subtracted from each main diagonal term).

Because the null space of a matrix is the set of solutions to the homogeneous system that has that matrix as coefficients matrix, we can also write $E_\lambda$ as the set of solutions of that system, which is the null space of the corresponding coefficients matrix as written above.

**PROBLEM.** You are told that $-1$ and $3$ are e-values of the matrix

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

Find a basis for each e-space $E_{-1}$ and $E_3$.

**PROBLEM.** You are told that $0$ and $2$ are e-values of the matrix

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

Find a basis for each e-space $E_0$ and $E_2$.

**PROBLEM.** You are told that $1$ is an e-value of the matrix

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

Find a basis for the e-space $E_1$.

**ALGORITHM (verifying if a scalar is an e-value)**

For a scalar $\lambda$ to be an e-value of a matrix $A$, there must exist vectors $u$ other than $0$ in $E_\lambda = \text{Null}(A - \lambda I)$

That is, it must be true that

$$\text{Nullity}(A - \lambda I) > 0$$

So we can check if $\lambda$ is or isn’t an e-vector by performing GE on $A - \lambda I$ to find its nullity.

- Apply this to show that the given e-values in the previous 3 examples really are e-values of their matrices.
- Apply this to show that $4$ is not an e-value of the matrix in the last example.
- Consider $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $Tu = 2u$. Let $A$ be the corresponding standard matrix.
  a) Without calculating $A$, what must be the only possible e-value $\lambda$ of $A$? What is the corresponding e-space?
  b) Now compute $A$ and use the algorithm to show that $\lambda$ really is an e-value.
  c) Use the algorithm to show that no other scalar other than $\lambda$ can be an e-value.
- Consider $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined as the rotation by $90^\circ$ counterclockwise. Let $A$ be the corresponding standard matrix.
  a) Without calculating $A$, explain why $A$ cannot have any e-vectors or e-values.
  b) Now compute $A$ and use the algorithm to show that no scalar $\lambda$ is an e-value.
An important reason to check whether 0 is an e-value:

**THEOREM (invertibility and zero e-value)**

A square matrix is invertible if and only if 0 is not an e-value.

*Proof.* A matrix is invertible if and only if the only solution to $Ax = 0$ is $x = 0$. But this system is the same as $Ax = \lambda x$ for the particular value $\lambda = 0$, and to say that this is only satisfied by the vector $x = 0$ means, by definition, that $\lambda = 0$ is not an e-value.

**THEOREM (e-values of powers and of inverse of a matrix)**

If a matrix $A$ has an e-vector $u$ corresponding to an e-value $\lambda$, then $u$ is also an e-vector of $A^k$ and of $A^{-1}$, but corresponding to e-values $\lambda^k$ and $1/\lambda$, respectively.

In particular, if the e-values of $A$ are

$$\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$$

then the e-values of $A^k$ are

$$\{\lambda_1^k, \lambda_2^k, \ldots, \lambda_p^k\}$$

and the e-values of $A^{-1}$ are

$$\left\{ \frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_p} \right\}$$

*Proof idea.* Suppose $\lambda$ is an eigenvalue of $A$ and $u$ is any corresponding e-vector. To check that $u$ is also an e-vector for the matrix $A^2$, but corresponding to $\lambda^2$, we need to show that $A^2u = \lambda^2 u$. And indeed:

$$A^2u = A(Au) = A(\lambda u) = \lambda(Au) = \lambda(\lambda u) = \lambda^2 u$$

This idea works for any power $k \geq 2$ (formally, it would be written as an induction proof).

As for $A^{-1}$, we must be able to prove that $A^{-1}u = (1/\lambda)u$. One possible way is to remember that $A^{-1}$ is the matrix of the inverse transformation associated to $A$. That transformation takes $u$ into $\lambda u$, so $A^{-1}$ must take $\lambda u$ into $u$:

$$A^{-1}(\lambda u) = u$$

Since $A^{-1}$ enjoys linearity properties, the scalar can come out:

$$\lambda(A^{-1}u) = u$$

And since we know $\lambda \neq 0$ (because we are assuming $A$ is invertible, so it can’t have 0 e-value), we can divide by it:

$$A^{-1}u = \frac{1}{\lambda}u$$

as we wanted.

**PROBLEM.** You are told that $\lambda = 5$ is an e-value of

$$A = \begin{bmatrix} 11 & 18 \\ -3 & -4 \end{bmatrix}$$

Compute $A^2$ and show that 25 is an e-value of it.

Finally, an important result that we’ll use in the near future (in the textbook, this is theorem 5.3, found in section 5.3, page 317):

**THEOREM (e-vectors from different e-spaces are independent)**

Let $A$ be a square matrix. If $u_1, u_2, \ldots, u_k$ are any number of e-vectors corresponding to different e-values, then they are linearly independent.
5.2. Characteristic polynomial

Key concepts: Characteristic polynomial, finding e-values, complex e-values and e-vectors

Some facts about polynomials

A polynomial of degree $m$ is an expression of the form

$$p(t) = a_m t^m + a_{m-1} t^{m-1} + \cdots + a_1 t + a_0$$

where $t$ is a variable and the $a_i$’s are scalars, with $a_m \neq 0$ (otherwise there wouldn’t be a $t^m$ term and we wouldn’t say that the degree is $m$).

Polynomials of degree $m$ can have at most $m$ real roots, that is, numbers $t$ such that $p(t) = 0$. If we allow complex numbers as possible roots, then we’ll always find exactly $m$ (but some might be repeated).

If $z_1, \ldots, z_k$ are the complex roots, each one repeated $n_1, \ldots, n_k$ times respectively, then the polynomial factors as

$$p(t) = a_m (t - z_1)^{n_1} (t - z_2)^{n_2} \cdots (t - z_k)^{n_k}$$

The numbers $n_i$ are called the multiplicity of their corresponding roots $z_i$.

The reason why we’re talking about polynomials is that we will associate to every square $m \times m$ matrix an $m$-degree polynomial that is closely connected to its e-values:

ALGORITHM (finding e-values)

The characteristic polynomial of a square matrix $A_{m \times m}$ is the $m$-degree polynomial

$$p(t) = \det(A - tI) = \begin{vmatrix} a_{11} - t & a_{12} & a_{13} & \cdots & a_{1m} \\ a_{21} & a_{22} - t & a_{23} & \cdots & a_{2m} \\ a_{31} & a_{32} & a_{33} - t & \cdots & a_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mm} - t \end{vmatrix}$$

Its real roots are the e-values of $A$, and the multiplicity of each real root is called the algebraic multiplicity of that e-value.

The book likes to call the algebraic multiplicity simply “multiplicity”.

Justification. A scalar $\lambda$ is an e-value of a matrix $A_{m \times m}$ if and only if $\text{nullity}(A - \lambda I) > 0$. Since $\text{rank}(A) + \text{nullity}(A) = m$, this is the same as $\text{rank}(A - \lambda I) < m$, and because $A - \lambda I$ is a square matrix, the only way that it can have a non-maximal rank like this is if $\det(A - \lambda I) = 0$.

- Find the e-values of the matrices

$$A = \begin{bmatrix} -4 & -3 \\ 3 & 6 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & -4 \end{bmatrix} \quad C = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

- If the characteristic polynomial of a $7 \times 7$ matrix is

$$p(t) = -(t + 2)^2(t - 3)(t + 1)^3(t - 5)$$

then what are its e-values and corresponding algebraic multiplicities?
• Find the e-values and corresponding e-vectors of the matrix

\[ D = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \]

(this example features some annoying irrational numbers).

• Show that the matrix

\[ D = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \]

doesn’t have any (real) e-values.

**E-values of triangular matrices**

The e-values of a triangular matrix (in particular also of a diagonal matrix) are the entries along the main diagonal.

In particular the same is true of a diagonal matrix.

*Justification.* Simply apply (to the matrix \( A - \lambda I \), which is also triangular) the fact that the determinant of a triangular matrix is the product of the entries on the main diagonal.

• What are the e-values of the following matrix?

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
2 & 3 & 0 & 0 \\
4 & 5 & 6 & 0 \\
7 & 8 & 9 & 10
\end{bmatrix}
\]

• What are the e-values and their algebraic multiplicities for the following matrix?

\[
\begin{bmatrix}
2 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 \\
0 & 0 & 2 & 0 \\
0 & 0 & 0 & 2
\end{bmatrix}
\]

**Row operations change eigenstuff!**

The characteristic polynomial of a matrix changes if we perform row operations on it. Therefore, e-values and e-vectors of a matrix might not be the same as those of its RREF, and we cannot perform GE to find e-values.

The matrix \( A = \begin{bmatrix} -4 & -3 \\ 3 & 6 \end{bmatrix} \) is invertible, therefore its RREF is \( R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \). But the e-values of \( A \) and \( R \) are not the same.

We’ve learned two different notions of *multiplicity* for e-values:

• *Geometric multiplicity* is the dimension of the eigenspace.

• *Algebraic multiplicity* is its multiplicity as a root of the characteristic polynomial.

A nontrivial relation exists between them (we won’t be concerned with a proof; even the textbook skips this particular proof):

**THEOREM (algebraic versus geometric multiplicity)**

The geometric multiplicity of an e-value is always less than or equal to the algebraic multiplicity.

Find all e-values and a basis for each e-space:

\[
\begin{bmatrix}
4 & 0 & 1 & 0 \\
0 & 3 & 0 & 0 \\
0 & 0 & 4 & 0 \\
0 & 0 & 0 & 3
\end{bmatrix}
\]
5.2. CHARACTERISTIC POLYNOMIAL

Geometric multiplicity is always at least 1; must be 1 if the algebraic multiplicity is 1

If λ is an e-value of a given matrix, its geometric multiplicity must be at least 1 (because being an e-value means that the e-space $E_\lambda$ is nontrivial).

If we compute the characteristic polynomial and discover that the algebraic multiplicity is 1, we can conclude by the previous theorem that $\dim(E_\lambda)$ must be 1.

PROBLEM. If the characteristic polynomial of a $6 \times 6$ matrix is

$$p(t) = (t - 1)^2(t + 1)(t - 2)(t + 2)^2$$

what can we conclude about the e-values and their geometric multiplicities?

We learned the concept of similar matrices in the last chapter. It turns out that any two matrices that are similar share the same e-values:

**THEOREM (e-values and eigenspaces of similar matrices)**

- If $A$ and $B$ are similar matrices, they have the same characteristic polynomial and in particular the same e-values with the same algebraic multiplicities.
- Furthermore, the geometric multiplicity of each e-value is also the same (but the e-vectors may be different).

*Proof sketch.* The characteristic polynomials of $A$ and $B$ are

$$p(t) = \det(A - tI) \quad \text{and} \quad q(t) = \det(B - tI)$$

Suppose there exists some invertible $P$ such that $B = P^{-1}AP$. The trick is to rewrite the $I$ in the definition of $q$ as $P^{-1}IP$, then factor the $P^{-1}$ on the left and the $P$ on the right, split the determinant, and cancel out the determinants of $P$ and $P^{-1}$ (which are numbers that are the inverse of each other):

$$q(t) = \det(B - tI) = \det(P^{-1}AP - tI)$$

$$= \det(P^{-1}AP - tP^{-1}IP)$$

$$= \det(P^{-1}(A - tI)P)$$

$$= \det(P)\det(A - tI)\det(P^{-1})$$

$$= \det(A - tI) = p(t)$$

proving that $p(t) = q(t)$. So the roots of these polynomials are the same, with the same multiplicities (algebraic multiplicity of the e-values).

The proof that the geometric multiplicities are also shared involves a trick that allows one to use $P$ to obtain a basis for $E_\lambda$ for the $B$ matrix from a basis for $E_\lambda$ for the $A$ matrix, for each e-value $\lambda$. We’re not gonna worry about that (see exercise 84 if you’re curious about it).

The significance of this theorem is that it allows us to think of the e-values and the characteristic polynomial of a linear operator as those of its matrix with respect to any basis, not just the standard matrix. Its matrices with respect to any bases are similar matrices, so the characteristic polynomial computed with any one of these matrices will always be the same.

PROBLEM. Compute the characteristic polynomial of the operator

$$T[x, y]^T = [x + y, x - y]^T$$

using its standard matrix, and using its matrix with respect to the basis

$$\{[-2, 7]^T, [-1, 3]^T\}$$

Now comes the only part of the entire course where we worry about complex numbers. Even though we only study real matrices and vectors, they can have complex e-values and e-vectors, because polynomials with real coefficients sometimes have complex roots. Complex e-values and e-vectors find applications even in some topics that are only concerned with real objects, most notably linear first-order systems of differential equations, also called continuous dynamical systems (Math 244, Math 252 at Rutgers), so it’s worth it to be able to work with them.
Complex eigenstuff

Any square matrix $A_{m \times m}$ has exactly $m$ e-values if they’re allowed to be complex numbers (counted with their algebraic multiplicity), because any $m$-degree polynomial has $m$ real and complex roots.

These e-values may only yield complex e-vectors, that is, $\dim(E_\lambda)$ as a real subspace will still be 0 if $\lambda$ is a complex e-value (when the entries of $A$ are all real).

The sum of all the algebraic multiplicities is $m$, the degree of the characteristic polynomial.

But the geometric multiplicities don’t need to add up to $m$.

If the entries of $A$ are all real and we find a complex e-value $\lambda$ with a complex e-value $u$, then the complex number $\lambda$ and the complex vector $\mathbf{u}$ obtained by flipping the sign of all instances of the imaginary number $i$ in them are also an e-value and an e-vector.

• Find the (complex) e-values and e-vectors of
  \[
  \begin{bmatrix}
  1 & -10 \\
  2 & 5
  \end{bmatrix}
  \]

• Find the (complex) e-values and e-vectors of the rotation matrix $R_{270^\circ}$.

Suggested problems from the book

• Systematic: 1-44
• True or false: 53-72
• Conceptual: 73, 76-78, 81, 83, 85-86

5.3. Diagonalization

Key concepts: diagonalizable matrices, diagonalization

**DEFINITION (diagonalizable matrix)**

A square matrix $A$ is called **diagonalizable** if it is similar to a diagonal matrix, that is, there exist some diagonal matrix $D$ and some invertible matrix $P$ such that

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

(in this topic, it’s more convenient to write this instead of the equivalent statement $A = Q^{-1}DQ$).

We already know that similar matrices share the same e-values, and that the e-values of a diagonal matrix are the entries along its diagonal. Therefore, when $A$ is diagonalizable, the only possible choice of $D$ matrix is a diagonal matrix having the e-values of $A$ along its diagonal, in some order.

• The matrix $A = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$ is diagonalizable with the choices
  \[
  D = \begin{bmatrix}
  2 & 0 \\
  0 & 4
  \end{bmatrix} \quad P = \begin{bmatrix}
  1 & -1 \\
  1 & 1
  \end{bmatrix}
  \]

  It is also diagonalizable with the choices
  \[
  D = \begin{bmatrix}
  4 & 0 \\
  0 & 2
  \end{bmatrix} \quad P = \begin{bmatrix}
  -2 & -5 \\
  2 & -5
  \end{bmatrix}
  \]

• The matrix $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ is diagonalizable with the choices
  \[
  D = \begin{bmatrix}
  1 & 0 & 0 \\
  0 & 2 & 0 \\
  0 & 0 & 3
  \end{bmatrix} \quad P = I_3
  \]

Similarly, any diagonal matrix is diagonalizable.
• The matrix \( C = \begin{bmatrix} 0 & 0 \\ 4 & 0 \end{bmatrix} \) is not diagonalizable. Its only e-value is 0 with algebraic multiplicity 2 (because it’s triangular with zeroes on the diagonal); therefore, if it were diagonal, the \( D \) matrix would have to be \( O \).

But then \( C \) would have been \( C = PDP^{-1} = POP^{-1} = O \) (the zero matrix), which it’s not.

**Motivation for diagonalization**

When we wish to think of square matrices \( A_{m \times m} \) as operators acting on \( \mathbb{R}^m \), the simplest ones to study are the diagonal matrices, which correspond to operators having very simple formulas in which the different entries of the vectors don’t mix together:

\[
X = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \quad \Rightarrow \quad Tx = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix}
\]

\[
Y = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \quad \Rightarrow \quad Ty = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 - x_2 \\ 3x_2 - x_1 \end{bmatrix}
\]

When we have a non-diagonal matrix, or some complicated operator corresponding to it, it might be that it would be diagonal in some other basis, so it would be more convenient to study it in that basis (study \([T]_{B,B}\) instead of the standard matrix \( A_T \)):

\[
B = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} \quad \Rightarrow \quad [Ty]_B = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}
\]

We already know that the standard matrix of an operator is related to its matrix in a different basis by similarity:

\( X = [T]_B = B^{-1}YB \quad \Rightarrow \quad Y = BXB^{-1} \)

Therefore, given a non-diagonal matrix, it’s helpful to know whether it is similar to a diagonal matrix, and if so how to find it.

**THEOREM (relation between diagonalization and eigenstuff)**

A matrix \( A_{m \times m} \) is diagonalizable if and only if there exists a basis for \( \mathbb{R}^m \) consisting of e-vectors of \( A \).

In that case, the matrices \( D \) and \( P \) such that \( A = PDP^{-1} \) can be taken as follows:

- \( P \) has independent e-vectors of \( A \) as columns.
- \( D \) has the corresponding e-values along the diagonal.

**PROBLEM.** You are told that the characteristic polynomial of

\[
A = \begin{bmatrix} -4 & -6 & 0 \\ 3 & 5 & 0 \\ 3 & 3 & 2 \end{bmatrix}
\]

is \( p(t) = -(t + 1)(t - 2)^2 \). Find a diagonalization of this matrix.

**PROBLEM.** Diagonalize the matrix

\[
A = \begin{bmatrix} 6 & 6 \\ -2 & -1 \end{bmatrix}
\]

**Justification for the diagonalization procedure.**

If \( A \) is diagonalizable, that is, \( A = PDP^{-1} \) for some invertible matrix \( P \) and diagonal matrix \( D \), let’s try to show that the columns \( p_1, p_2, \ldots, p_m \) of \( P \) are e-vectors of \( A \) corresponding to the entries \( d_1, d_2, \ldots, d_m \) of \( D \) as e-values. Remember that:

- the columns of a matrix-matrix product \( BC \) are the matrix-vector product between \( B \) and the columns of \( C \);
- the multiplication with a diagonal matrix on the right causes the columns to be multiplied by the entries of that diagonal matrix.
Then:
\[
\begin{bmatrix}
A p_1 & A p_2 & \cdots & A p_m
\end{bmatrix} = AP = (PDP^{-1})P = PD = \begin{bmatrix}
d_1 p_1 & d_2 p_2 & \cdots & d_m p_m
\end{bmatrix}
\]
This shows that \( A p_i = d_i p_i \), as we wanted.

Reciprocally, suppose that there exists a basis \( B = \{u_1, u_2, \ldots, u_m\} \) for \( \mathbb{R}^m \) satisfying \( A u_i = \lambda_i u_i \) for some \( \lambda_i \) \( (i = 1, 2, \ldots, m) \). Let’s try to show \( A \) is diagonalizable. Consider the linear operator \( T_A \) related to \( A \). By definition, its matrix in the basis \( B \), denoted \( [T_A]_B \), is diagonal with the \( \lambda_i \)'s as entries, because the way to write each \( T_A u_i \) in the basis \( B \) is with a bunch of zeroes and a \( \lambda_i \):
\[
T_A u_i = 0 u_1 + 0 u_2 + \cdots + 0 u_{i-1} + \lambda_i u_i + 0 u_{i+1} + \cdots + 0 u_m
\]
But we also know of a formula to compute \( [T_A]_B \):
\[
[T_A]_B = B^{-1} A B
\]
where \( B \), the matrix associated to the basis \( B \), has the basis vectors as columns:
\[
B = \begin{bmatrix}
u_1 & u_2 & \cdots & u_m
\end{bmatrix}
\]
Multiply both sides on the left by \( B \) and on the right by \( B^{-1} \) to obtain a diagonalization for \( A \), with \( P = B \) and \( \textbf{D} = [T_A]_B \).

This algorithm is used to find a diagonalization when it exists. The next theorem, whose proof we omit, tells us when that happens:

**THEOREM (possibility of diagonalization is determined by multiplicities)**

An \( m \)-dimensional square matrix is diagonalizable if and only if both of these conditions are true:

- The sum of the algebraic multiplicities of the \( \lambda \)-values is \( m \).
- Each geometric multiplicity is equal to the corresponding algebraic multiplicity.

In particular, if all \( \lambda \)-values are distinct (that is, they are all real and each has algebraic multiplicity 1), the matrix is diagonalizable. This is because, since each algebraic multiplicity is 1, so must be each geometric multiplicity.

- The matrix
  \[
  \begin{bmatrix}
  4 & 0 & 1 & 0 \\
  0 & 3 & 0 & 0 \\
  0 & 0 & 4 & 0 \\
  0 & 0 & 0 & 3
  \end{bmatrix}
  \]
  is **not** diagonalizable, because the algebraic multiplicity of \( \lambda = 4 \) is 2, but we’ve determined previously that \( \text{dim}(E_4) = 1 \).
- A \( 3 \times 3 \) matrix with characteristic polynomial
  \[
p(t) = -(t + 2)(t^2 + 9)
  \]
cannot be diagonalizable. Its only \( \lambda \)-value is \(-2\), with algebraic multiplicity 1, but the sum of all algebraic multiplicities would need to be 3.
- A \( 4 \times 4 \) matrix with characteristic polynomial
  \[
p(t) = t(t - 3)(t + 2)(t + 1)
  \]
is diagonalizable, since all \( \lambda \)-values are distinct.
- A \( 4 \times 4 \) matrix with characteristic polynomial
  \[
p(t) = (t + 1)(t + 2)^2(t + 3)
  \]
is diagonalizable if and only if the \( \lambda \)-space \( E_{-2} \) has dimension 2. For example, \( A \) below is diagonalizable (it is even diagonal), but \( B \) is not (check):
\[
A = \begin{bmatrix}
-1 & 0 & 0 & 0 \\
0 & -2 & 0 & 0 \\
0 & 0 & -2 & 0 \\
0 & 0 & 0 & -3
\end{bmatrix}
\quad B = \begin{bmatrix}
-1 & 0 & 0 & 0 \\
0 & -2 & 5 & 0 \\
0 & 0 & -2 & 0 \\
0 & 0 & 0 & -3
\end{bmatrix}
\]
5.4. Diagonalization of operators

Key concepts: diagonalization applied to linear operators

The book tries too hard to differentiate matrices from the linear transformations associated to them. This section is simply about applying the same eigen-ideas studied so far to linear operators, instead of the matrices that represent them.

PROBLEM. Find, if possible, a basis of e-vectors for the operator

\[ T[x, y, z]^T = [x + 2y + z, 2y, -x + 2y + 3z]^T \]

PROBLEM. Find, if possible, a basis of e-vectors for the operator

\[ T[x, y]^T = [-7x - 10y, 3x + 4y]^T \]

5.5. Applications of diagonalization

Key concepts: computing powers of a matrix, Markov Chains, difference equations

Knowing how to diagonalize of a square matrix allows us to easily compute any power of it. The textbook doesn’t make explicit mention of this important property, but it uses it often and we should be aware of it:

**ALGORITHM** (computing powers of diagonal and diagonalizable matrices)

- To compute powers of a diagonal matrix, we can just compute the powers of each entry:

\[
D = \begin{bmatrix}
d_1 & 0 & 0 & \cdots & 0 \\
0 & d_2 & 0 & \cdots & 0 \\
0 & 0 & d_3 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & d_m
\end{bmatrix} \quad \Rightarrow \quad D^k = \begin{bmatrix}
d_1^k & 0 & 0 & \cdots & 0 \\
0 & d_2^k & 0 & \cdots & 0 \\
0 & 0 & d_3^k & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & d_m^k
\end{bmatrix}
\]

- If \( A \) is a diagonalizable matrix: \( A = PDP^{-1} \) with \( D \) being diagonal, then the powers of \( A \) are given by

\[ A^k = PD^k P^{-1} \]

*Justification.* The first assertion is seen by direct computation of matrix products. The second assertion is due to the associativity of matrix product: when we multiply \( A \) by itself any number of times and rewrite each
copy of $A$ as $PDP^{-1}$, the $P$'s in the middle go away. For example:

$$A^4 = AAAAA = (PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1})$$

$$= PD(PP^{-1})D(PP^{-1})D(PP^{-1})DP^{-1}$$

$$= PDDDP^{-1}$$

$$= PD^3P^{-1}$$

- Compute $A^2$ and $A^3$:

$$A = \begin{bmatrix} 16 & -3 \\ 90 & -17 \end{bmatrix}$$

Also compute $A^2$ directly to check. The diagonalization is

$$A = \begin{bmatrix} 1 & -1 \\ 6 & -5 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -5 & 1 \\ -6 & 1 \end{bmatrix}$$

Now we turn to an important topic in probability theory.

**Motivation (Markov Chains)**

Suppose that it is raining today and that:

- If it rains in a day, there’s 20% chance it will rain the next day.
- If it doesn’t rain in a day, there’s 50% chance it will rain the next day.

Say that, on a given day, we are in state $R$ if it rains and state $NR$ if it doesn’t. We change state randomly every day, and what determines the probabilities of the next day’s state is the present day’s state. We can also say that our initial state is $R$ with 100% probability, so the next day’s probabilities are 20% for $R$ and 80% for $NR$.

To compute the probability of $R$ two days from now, we compound probabilities: it will be 20% in case the 20% chance of $R$ tomorrow materializes, and 50% in case the 80% chance of $NR$ tomorrow becomes true. Similarly for the probability of $NR$:

$$P(R) \text{ (in 2 days)} = 0.2 \times 0.2 + 0.8 \times 0.5 = 0.04 + 0.4 = 0.44$$

$$P(NR) \text{ (in 2 days)} = 0.2 \times 0.8 + 0.8 \times 0.5 = 0.16 + 0.4 = 0.56$$

Now we can use these numbers to find the probabilities for the following day, and so on and so on:

$$P(R) \text{ (in 3 days)} = 0.44 \times 0.2 + 0.56 \times 0.5 = 0.088 + 0.28 = 0.368$$

$$P(NR) \text{ (in 3 days)} = 0.44 \times 0.8 + 0.56 \times 0.5 = 0.352 + 0.28 = 0.632$$

$$P(R) \text{ (in 4 days)} = 0.368 \times 0.2 + 0.632 \times 0.5 = 0.0736 + 0.316 = 0.3896$$

$$P(NR) \text{ (in 4 days)} = 0.368 \times 0.8 + 0.632 \times 0.5 = 0.2944 + 0.316 = 0.6104$$

Note how these probabilities seem to be converging. It is expected that they should converge to some distribution where the probability of $R$ is less than that of $NR$, no matter where we started (the probability distribution today). Indeed, because the hypotheses say that rain today implies a small chance of rain tomorrow, we should expect that the probability of $R$ on some day in the distant future is smaller than that of $NR$; and as we calculate further and further days, the uncertainties accumulate too much for the initial condition to have any effect.

Define 2D vectors $u_n$, for $n = 0, 1, 2, \ldots$, containing the probabilities of $R$ and $NR$ for each day $n$ (we call the starting day 0). We know

$$u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

To obtain the probabilities for day $n + 1$, we compound them with day $n$ probabilities the same way we did for 2 days from now: if $u_n = [p, q]^T$ then

$$u_{n+1} = \begin{bmatrix} 0.2p + 0.5q \\ 0.8p + 0.5q \end{bmatrix} = \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix} u_n$$

The matrix $A = \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}$ contains the probabilities of moving from a state (indexed in the column) to another (indexed in the row). For example, it has 80% in column 1 and row 2, which is the probability of transitioning from state 1 ($R$) to state 2 ($NR$). Note that the sum of all entries in any column is 1, because they are complementary probabilities.
To obtain $u_n$ from $u_0$, we need to multiply $n$ times by $A$:

$$u_n = A^n u_0$$

So it will be useful to have a diagonalization of $A$ to easily compute its powers.

Another question we might ask is the following: is there a stationary distribution of probabilities? That is, an assignment of probabilities for $R$ and $NR$ in a given day such that the probabilities computed for the next day from it would not change? In other words, we want to find a vector $v$ (the sum of whose entries is 1, because they represent complementary probabilities) such that

$$Av = v$$

which is to say an e-vector of $A$ with e-value 1. With a diagonalization of $A$, we can check if it has e-value 1 (it does), and once we find some corresponding e-vector, we can always consider an appropriate multiple of it to find another e-vector whose entries add up to 1.

**DEFINITION (Markov Chains)**

A Markov Chain is a random process involving some object that at each time can be in one of a few possible states, and such that each of the probabilities of it passing from any given state to any other given state are fixed throughout the process.

**THEOREM (Markov Chains)**

Suppose a Markov Chain has $m$ possible states $S_1, \ldots, S_m$. Consider the stochastic matrix $A_{m \times m}$ defined as follows:

$$a_{ij} = \text{Probability of passing from state } S_j \text{ to state } S_i \quad \text{(note the order!)}$$

If $u_n$ is an $m$-dimensional vector containing the probabilities of being in each state at step $n$, with $u_0$ containing the initial probabilities, then

$$u_n = A^n u_0$$

The matrix $A$ has e-value 1. In many cases (some technical assumptions are necessary), the dimension of the corresponding e-space will be 1, so there will always be exactly one stationary probability vector, that is, a vector $v$ with the following properties:

- $Av = v$
- $v_1 + v_2 + \cdots + v_m = 1$

Furthermore, again in many cases, for any initial vector $u_0$ (the sum of whose entries is 1) the sequence of vectors $u_n$ converges to $v$.

Idea of proof. The sum of the entries in each column of $A$ is 1. That means the sum of each column in $A - I$ is zero (there’s a 1 subtracted from exactly one entry in each column). So all columns of $A - I$ satisfy the same linear equation of having the sum of their entries zero, which makes them linearly dependent. That means $\det(A - I) = 0$, that is, 1 is an e-value.

Proving that its geometric multiplicity is 1 when $A$ is regular enough is trickier (we’d have to identify, first of all, what we mean by regular enough). Proving that any initial probability vector converges to the only stationary probability vector involves proving that the other e-values are all strictly less than 1 in absolute value, which means their powers converge to zero. Knowing the size of these other e-values even gives an estimate on how fast the chain converges.

- Find the unique stationary probability vector in the R/NR example above. Also find the transition matrix for 5 steps.
- Find whether there is a unique stationary probability vector for the Markov Chain having 3 states $A, B, C$ with the following transition probabilities:

  From $A$ to $B$ : 0.2  From $A$ to $C$ : 0.1  
  From $B$ to $A$ : 0.4  From $B$ to $C$ : 0.0  
  From $C$ to $A$ : 0.0  From $C$ to $B$ : 1

- Google Page Rank utilizes (among other things) an enormous Markov Chain where each website is a state.
Now for a different topic where diagonalization of matrices is helpful:

Motivation (difference equations)

You’ve probably heard of the Fibonacci sequence:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \ldots$$

This is the sequence, starting from 0, 1, whose elements are obtained consecutively by adding the 2 previous elements. If we denote its general element by $F_n$, where the index $n$ starts at 0, we have the defining recursive rule (or difference equation) and initial conditions:

$$\begin{cases} 
F_n = F_{n-1} + F_{n-2} & \text{for all } n \geq 2 \\
F_0 = 0, \ F_1 = 1 
\end{cases}$$

This sequence and other related sequences appear in many different areas of pure and applied Math. It is connected to the golden ratio, that is, the number

$$\varphi = \frac{1 + \sqrt{5}}{2} = 1.618033 \ldots$$

by the fact that the ratio of consecutive elements of the sequence gets closer and closer to it:

$$\lim_{n \to \infty} \frac{F_{n+1}}{F_n} = \varphi$$

What you may not know is that there is a closed-form formula for the general element in this sequence, which surprisingly involves irrational numbers:

$$F_n = \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left( \frac{1 - \sqrt{5}}{2} \right)^n$$

Expanding the terms in parentheses by the binomial theorem, all even powers of $\sqrt{5}$ cancel out, leaving only odd powers, which divided by $\sqrt{5}$ make integers:

$$\begin{align*}
F_0 &= \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^0 - \frac{1}{\sqrt{5}} \left( \frac{1 - \sqrt{5}}{2} \right)^0 = \frac{1}{\sqrt{5}} (1 - 1) = 0 \\
F_1 &= \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^1 - \frac{1}{\sqrt{5}} \left( \frac{1 - \sqrt{5}}{2} \right)^1 = \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} - \frac{1 - \sqrt{5}}{2} \right) = \frac{1}{\sqrt{5}} \sqrt{5} = 1 \\
F_2 &= \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^2 - \frac{1}{\sqrt{5}} \left( \frac{1 - \sqrt{5}}{2} \right)^2 = \frac{1}{\sqrt{5}} \left( \frac{1 + 2\sqrt{5} + 5}{4} - \frac{1 - 2\sqrt{5} + 5}{4} \right) = \frac{1}{\sqrt{5}} \frac{4\sqrt{5}}{4} = 1 
\end{align*}$$

This closed-form formula allows us to compute or approximate the value of $F_n$ for large $n$ without needing to compute all terms recursively up to $n$, and it proves the statement about the golden ratio too (a simple Calculus exercise).

It turns out that matrices and their eigenstuff provide a way to find this unexpected closed-form formula. The trick is that the formula

$$F_{n+2} = F_{n+1} + F_n$$

can be written as one of the two equations of the following matrix equation (the other being the trivial $F_{n+1} = F_{n+1}$):

$$\begin{bmatrix} F_{n+1} \\ F_{n+2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix}$$

Call $\mathbf{u}_n$ a vector containing both $F_n$ and $F_{n+1}$:

$$\mathbf{u}_n = \begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix}$$

So $\mathbf{u}_{n+1}$ is obtained from the previous $\mathbf{u}_n$ by multiplication with a simple square matrix. The initial conditions give us the first of the $\mathbf{u}_n$ vectors:

$$\mathbf{u}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
To obtain \( u_1 \), we would compute
\[
\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\]

To obtain \( u_2 \), we would compute
\[
\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}
\]

To obtain \( u_n \), instead of doing this \( n \) times, we can realize that we just need to compute powers of the matrix:
\[
\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} u_{n-1} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \cdots = \begin{bmatrix} 0 \\ 1 \end{bmatrix}^n u_0 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\]

We know that powers can be easily computed if we have a diagonalization. For this particular matrix, it turns out that we can take
\[
\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 1 + \sqrt{5} \\ 1 & 1 - \sqrt{5} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{5} & 1 \\ 1/\sqrt{5} & -1 \end{bmatrix}
\]

We don’t care about the second entry, because the first one is the one that contains \( F_n \).

We could contemplate sequences where there are \( k \) initial conditions instead of 2, and then every term after them is obtained by summing the previous \( k \) terms. Or instead of summing, it could be some linear combination of them:

**DEFINITION (recursive sequence and difference equation)**

A recursive sequence of degree \( k \) is a sequence \( a_n \), indexed for \( n \geq 0 \), whose terms satisfy a relation of the form below (called a difference equation):
\[
a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k} \quad \text{for all } n \geq k
\]

for some fixed real numbers \( c_1, \ldots, c_k \).

It is completely determined once we fix the first \( k \) terms, called the **initial conditions**:
\[
a_0, a_1, \ldots, a_{k-1}
\]

- The sequence that starts out as \( a_0 = 1, \ a_1 = 2, \ a_2 = 3 \) and is such that every entry from there on is the arithmetic mean of the 3 previous entries is a recursive sequence with difference equation
\[
a_n = \frac{1}{3} a_{n-1} + \frac{1}{3} a_{n-2} + \frac{1}{3} a_{n-3}
\]

It starts
\[
1, 2, 3, 2, \frac{7}{3}, \frac{22}{9}, \frac{61}{27}, \ldots
\]

- The sequence of the powers of 2
\[
1, 2, 4, 8, 16, 32, 64, \ldots
\]
is a degree 1 recursive sequence:

\[ a_0 = 0, \quad a_n = 2a_{n-1} \]

The formula for the general term is

\[ a_n = 2^n \]

**ALGORITHM (finding the general term of a recursive sequence)**

Suppose a sequence \( a_n \) satisfies a difference equation of degree \( k \):

\[ a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_k a_{n-k} \quad \text{for all } n \geq k \]

Denote by \( u_0 \) the vector containing the \( k \) initial conditions:

\[ u_0 = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{k-1} \end{bmatrix}^T \]

and by \( u_n \) the \( k \)-dimensional vectors containing consecutive entries starting at \( a_n \):

\[ u_n = \begin{bmatrix} a_n & a_{n+1} & a_{n+2} & \cdots & a_{n+k} \end{bmatrix}^T \]

Then we have \( u_n = A^n u_0 \), where \( A \) is the matrix

\[
A = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 1 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 1 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & 1 \\
c_k & c_{k-1} & c_{k-2} & \cdots & c_3 & c_2 & c_1 \\
\end{bmatrix}
\]

(it has the coefficients \( c_i \) on the last row, in reverse order, and the diagonal above the main diagonal filled with 1's).

To compute the general element \( a_n \), compute the first entry of \( u_n \) using eigenstuff to raise \( A \) to \( n \).

- Compute a few of the terms of the recursive sequence with initial conditions \( a_0 = 0, \ a_1 = 1 \) and difference equation

\[ a_n = 3a_{n-1} - 2a_{n-2} \]

By inspection, try to guess what the general formula is, then find it using eigenstuff.

- What is the degree of the difference equation below?

\[ a_n = 7a_{n-2} + 6a_{n-3} \]

What is the general term of the sequence that starts \( a_0 = a_1 = a_2 = 1 \) and follows this equation? You can use the fact that

\[ -t^3 + 7t + 6 = -(t + 1)(t + 2)(t - 3) \]

The diagonalization needed in this question is:

\[
\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & 7 & 0 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & 3 \\ 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 30 & 5 & -5 \\ -12 & -8 & 4 \\ 2 & 3 & 1 \end{bmatrix}
\]

**Suggested problems from the book**

- **Systematic**: 57-62 (in section 5.3), 21-32, 70-79
- **True or false**: 1, 3-5, 12
Orthogonality

Orthogonal is the same as perpendicular. Geometry is much easier to study in terms of orthogonal frames of reference (we are much too used to thinking of 3D space in terms of length, width, height). Given two vectors of any dimension, a simple calculation can tell us whether they are orthogonal to each other. We’ll study ways to determine when a set of vectors are mutually orthogonal, and develop an algorithm (the Gram-Schmidt algorithm) that produces orthogonal vectors from any given initial vectors preserving their span. The orthogonal projection of a point \( P \) onto the subspace \( W \) generated by those vectors (that is, the “shadow” that it casts there) can then be easily computed, and it turns out to be the closest point to \( P \) in \( W \). An unexpected application of this property is the Least Squares algorithm, that is able to produce the best-fitting curve of a given type for a given collection of data points. Then we’ll study the geometric properties of orthogonal matrices (those whose columns are orthonormal vectors) and the orthogonal diagonalization properties of symmetric matrices (those equal to their own transpose).

6.1. Geometry of vectors

Key concepts: norm, distance, dot product, orthogonality, orthogonal projection onto a line

<table>
<thead>
<tr>
<th>Motivation for dot products and norm</th>
</tr>
</thead>
<tbody>
<tr>
<td>When are two vectors in ( \mathbb{R}^2 ) perpendicular to each other?</td>
</tr>
<tr>
<td>If the vectors correspond to points ((a, b)) and ((c, d)), the equation of the lines spanned by them are (rise over run)</td>
</tr>
<tr>
<td>( y = \frac{b}{a} x ) ( y = \frac{d}{c} x )</td>
</tr>
<tr>
<td>(only when ( a, c \neq 0 )). You probably remember that two lines are perpendicular when their slopes are the opposite and inverse of each other:</td>
</tr>
<tr>
<td>( \frac{b}{a} = -\frac{1}{\frac{d}{c}} = -\frac{c}{d} )</td>
</tr>
<tr>
<td>Multiplying out and moving everything to the left, this implies</td>
</tr>
<tr>
<td>( ac + bd = 0 )</td>
</tr>
</tbody>
</table>

Ideas like this (with some careful considerations of what happens if \( a = 0 \) or \( c = 0 \)) can be used to show that two vectors \( \mathbf{u} = [u_1, u_2]^T \) and \( \mathbf{v} = [v_1, v_2]^T \) in \( \mathbb{R}^2 \) are perpendicular if and only if the quantity

\[
u_1 v_1 + u_2 v_2
\]

(formed by products of the corresponding entries in each vector) is zero. We give this quantity the symbol \( \mathbf{u} \cdot \mathbf{v} \).

Now consider the expression \( \mathbf{u} \cdot \mathbf{u} \), that is, the quantity above formed by a vector and itself:

\[
\mathbf{u} \cdot \mathbf{u} = u_1^2 + u_2^2
\]

This is a sum of two squares, so it will only be zero when \( \mathbf{u} = [0, 0]^T = \mathbf{0} \) (makes sense: nonzero vectors are not perpendicular to themselves). But the geometric significance of this expression comes from the Pythagorean Theorem:

Therefore \( \mathbf{u} \cdot \mathbf{u} \) is the length-squared of the arrow that represents \( \mathbf{u} \).
ORTHOGONALITY

**Norm** of a vector: The length of the arrow that represents it in \( \mathbb{R}^m \):

\[
\|u\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_m^2}
\]

**Unit vector**: A vector with norm 1. If \( u \neq 0 \) is any nonzero vector, we can find a unit vector that points in the same direction as \( u \) by dividing it by its norm (that is, multiplying by the scalar which is the inverse of its norm):

\[
\frac{1}{\|u\|} u \text{ is a unit vector}
\]

**Distance** between two points: The length of the arrow connecting these points. If we think of the points as vectors, this arrow is equal to one vector minus the other, so the distance is:

\[
dist(P,Q) = \|u - v\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_m - v_m)^2}
\]

**Dot product** between two vectors: A scalar associated to them that has applications to the measurement of angles (studied in Calc 3):

\[
\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_mv_m
\]

This is not the same as the multiplication of the vectors as matrices, which would be written \( \mathbf{uv} \) without a dot (but would be undefined unless \( m = 1 \)).

- Find the norm of \([3, 0, -2]^T\).
- Find the distance between \([1, 2, 3]^T\) and \([-1, 2, 4]^T\).
- Find the dot product between the two vectors in the previous item.

**Dot product as a matrix product**

The dot product \( \mathbf{u} \cdot \mathbf{v} \) between two vectors (to be more precise, a \( 1 \times 1 \) matrix containing its value) actually can be represented as a matrix product: between the row vector \( \mathbf{u} \) and the column vector \( \mathbf{v} \):

\[
\mathbf{u}^T \mathbf{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix} = [u_1v_1 + u_2v_2 + \cdots + u_mv_m] = [\mathbf{u} \cdot \mathbf{v}]
\]

This is sometimes useful to prove theorems about dot products.

**Properties** (straightforward properties of norms, dot product and orthogonality)

*See also Theorem 6.1 on page 364.*

Let \( \mathbf{u}, \mathbf{v}, \mathbf{w} \) be vectors in \( \mathbb{R}^m \), let \( r \) be a scalar. Then:

- Dot product with itself gives norm square: \( \mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2 \)
- \( \mathbf{0} \) is the only vector with norm 0: if \( \|\mathbf{u}\| = 0 \) then \( \mathbf{u} = \mathbf{0} \)
- Commutativity: \( \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \)
- Distributive: \( \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} \)
- Associativity with scalar product: \( \mathbf{u} \cdot (r\mathbf{v}) = (r\mathbf{u}) \cdot \mathbf{v} = r(\mathbf{u} \cdot \mathbf{v}) \)
- Homogeneity of the norm: \( \|r\mathbf{u}\| = |r|\|\mathbf{u}\| \) (a scalar comes out of the norm with an absolute value)

Suppose that \( \|\mathbf{u}\| = 2, \|\mathbf{v}\| = 3 \) and \( \mathbf{u} \cdot \mathbf{v} = -4 \). Find \( \|5\mathbf{u} - 7\mathbf{v}\| \).
PROPERTIES (non-obvious properties of norms, dot product and orthogonality)

See pages 368 and 369.

Let \( \mathbf{u}, \mathbf{v} \) be vectors in \( \mathbb{R}^m \). Then:

- **Relation to matrix transpose:** \( (A\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (A^T \mathbf{v}) \) for any matrix \( A_{m \times m} \)
- **Cauchy-Schwarz Inequality:** \( |\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \|\mathbf{v}\| \)
- **Triangle Inequality:** \( \|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \)
- Dot product and orthogonality: Vectors \( \mathbf{u} \) and \( \mathbf{v} \) are **orthogonal** (perpendicular) if and only if \( \mathbf{u} \cdot \mathbf{v} = 0 \) (The last property is the main use for the dot product).

**Proof ideas.**

The first item is most easily proved using the expression of the dot product as a matrix product, and then using properties of matrix transpose and multiplication:

\[
(A\mathbf{u}) \cdot \mathbf{v} = (A\mathbf{u})^T \mathbf{v} = (\mathbf{u}^T A^T) \mathbf{v} = \mathbf{u}^T (A^T \mathbf{v}) = \mathbf{u} \cdot (A^T \mathbf{v})
\]

The Cauchy-Schwarz inequality is a nontrivial property that can be proved exploiting the following idea: consider a general scalar \( t \) and the expression

\[
f(t) = \|\mathbf{u} + t\mathbf{v}\|^2 \geq 0
\]

We can compute \( f(t) \) using the dot product (the first item) and distribute everything out to get a quadratic polynomial in \( t \). The fact that \( f(t) \geq 0 \) for all \( t \in \mathbb{R} \) puts restrictions on the coefficients of this polynomial (use the quadratic formula) that, when simplified, give the statement that \( |\mathbf{u} \cdot \mathbf{v}| - \|\mathbf{u}\| \|\mathbf{v}\| \leq 0 \).

To prove the Triangle Inequality, use dot products, distribute out, use the Cauchy-Schwarz inequality, and then the binomial theorem:

\[
\|\mathbf{u} + \mathbf{v}\|^2 = (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} = \|\mathbf{u}\|^2 + 2\mathbf{u} \cdot \mathbf{v} + \|\mathbf{v}\|^2 \\
\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\| \|\mathbf{v}\| + \|\mathbf{v}\|^2 \\
\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\| \|\mathbf{v}\| + \|\mathbf{v}\|^2 = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2
\]

Taking square roots gives the inequality.

The last item requires the same sort of reasoning that we did for 2 vectors on the plane. It’s possible to reduce it to that case because, even though we may be in a general \( \mathbb{R}^m \) space, we are still working with just two vectors and we can restrict our attention to the plane that they generate.

- Verify the first item for some \( A, \mathbf{u}, \mathbf{v} \).
- Verify Cauchy-Schwarz for \( [1, -1]^T \) and \( [-3, 4]^T \).
- Verify triangle inequality for \( [2, 3]^T \) and \( [4, 4]^T \).
- Which of the following pairs of vectors are orthogonal?
  - \( [2, 3, 0]^T \) and \( [6, -4, 1]^T \)
  - \( [1, 2, 3]^T \) and \( [-4, 3, -1]^T \)
- Which pairs among the following vectors are orthogonal?
  - \( [-2, -5, 3]^T \) , \( [1, -1, 2]^T \) , \( [-3, 1, 2]^T \)
- The vector \( \mathbf{0} \) is orthogonal to any other vector.
ORTHOGONALITY

DEFINITION (orthogonal projection on a line)
Given a line \( L \) and a point \( P \not\in L \), both in \( \mathbb{R}^m \), the orthogonal projection or simply projection of \( P \) onto \( L \) is the point \( P_0 \in L \) that is the closest to \( P \).

\( P_0 \) is like the shadow cast by \( P \) onto \( L \) through a direct (perpendicular to \( L \)) beam of light, that is, the line \( \overrightarrow{P_0P} \) is perpendicular to \( L \).

If \( P \in L \) to begin with, the projection is defined as \( P_0 = P \).

THEOREM (calculation of projection on a line through the origin)
Let \( u, v \in \mathbb{R}^m \) with \( v \neq 0 \). Denote by \( P \) the point represented by \( u \) and by \( L \) the line obtained extending the arrow for \( v \). Denote by \( w \) the vector represented by the orthogonal projection \( P_0 \) of \( P \) onto \( L \). Then

\[
w = \left( \frac{u \cdot v}{\|v\|^2} \right) v \quad \text{and} \quad \text{dist}(P, P_0) = \left\| u - \left( \frac{u \cdot v}{\|v\|^2} \right) v \right\|
\]

Proof. The formula for the distance is simply the calculation of \( \|u - w\| \). What we really need to prove is why \( w \) is given as in the formula above.

The geometric thinking suggests that \( P_0 \) is determined by the property that the vector \( \overrightarrow{P_0P} = u - w \) is orthogonal to \( v \). So the vector \( w \) that we are trying to find is the only vector on the line spanned by \( v \) (that is, \( w = cv \) for some constant \( c \)) such that \( u - w \) is orthogonal to \( v \):

\[
(u - cv) \cdot v = 0
\]

Use the properties of dot products to find \( c \) from this equation and you’ll see that it must be as in the theorem.

- Calculate (and draw) the projection of \( P = (0, 5) \) onto the line \( y = \frac{1}{2}x \). Calculate the distance from \( P \) to that line.
- Calculate (and draw) the projection of \( u = [1, 2, 3]^T \) onto the lines spanned by the following vectors:
  \[
v = [0, 1, 0]^T \quad \text{and} \quad w = [-1, -2, -3]^T
\]

REMARK (projection onto a line that doesn’t go through the origin)
The points on a line that goes through the origin are parametrized as \( cv \), where \( v \) is the vector giving the direction of the line. The points on a line not going through the origin, but instead going through some point \( P \), are parametrized as \( P + cv \). It’s possible to use this to obtain a formula for the projection onto a line that doesn’t pass through the origin. It’s just a small modification of the case presented above and we won’t consider it in this class.

Decomposition of a vector into projected and orthogonal parts with respect to another vector
Let \( u, v \in \mathbb{R}^m \) with \( v \neq 0 \). Then \( u \) can be written in a unique way as the sum \( u_1 + u_2 \) of a vector \( u_1 \) parallel to \( v \) and a vector \( u_2 \) orthogonal to \( v \):

They are calculated as, respectively:
• The projection onto the line spanned by \( v \).
• The difference between \( u \) and this projection (so that the sum of the two is \( u \)).

\[
\mathbf{u}_1 = \left( \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{v}||^2} \right) \mathbf{v} \quad \text{and} \quad \mathbf{u}_2 = \mathbf{u} - \mathbf{u}_1
\]

**Justification.** That these \( u_1, u_2 \) are parallel and orthogonal to \( v \) can be seen easily. To see they are unique, suppose that there could be some other two vectors \( w_1 \) and \( w_2 \) with the same properties:

• \( w_1 \) is parallel to \( v \)
• \( w_2 \) is orthogonal to \( v \)
• \( u = w_1 + w_2 \)

Then \( u_1 + u_2 = u = w_1 + w_2 \), which implies

\[
\mathbf{u}_1 - \mathbf{w}_1 = \mathbf{w}_2 - \mathbf{u}_2
\]

On this equation, the left side is parallel to \( v \) (sum of two vectors parallel to \( v \)) and the right side is orthogonal to \( v \) (take the dot product with \( v \) to see it’s zero). So the vector \( u_1 - w_1 = w_2 - u_2 \) is both parallel and orthogonal to \( v \). Since \( v \neq 0 \), this is only possible if it is \( 0 \), that is,

\[
\mathbf{u}_1 = \mathbf{w}_1 \quad \text{and} \quad \mathbf{u}_2 = \mathbf{w}_2
\]

proving that \( u_1, u_2 \) are unique.

• Write \( \mathbf{u} = [-3, 2]^T \) as the sum of a vector parallel to and a vector orthogonal to the vector \( \mathbf{v} = [1, 0]^T \). Draw them.
• Do the same for \( \mathbf{u} = [0, 1]^T \) and \( \mathbf{v} = [1, 1]^T \).

**Suggested problems from the book**

• Systematic: 1-16, 25-60
• True or false: 61-80
• Conceptual: 86, 91-92, 95

---

### 6.2. Orthogonal vectors

**Key concepts:** orthogonal and orthonormal sets and bases, Gram-Schmidt algorithm

**DEFINITION (orthogonal and orthonormal sets)**

Orthogonal set: A set of vectors in some Euclidean space that are all orthogonal to each other.

Orthonormal set: An orthogonal set with the additional property that each vector has norm 1.

Given an orthogonal set that doesn’t include \( 0 \), we can find an orthonormal set that has the same span as the span of the original set simply by normalizing the vectors, that is, dividing them by their norm to obtain vectors of norm 1.

Orthogonal / Orthonormal basis: A basis (for some subspace \( W \)) that is also an orthogonal / orthonormal set.

• The set \( \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m\} \subseteq \mathbb{R}^m \) is an orthonormal basis for \( \mathbb{R}^m \). The set containing only some of the standard vectors is an orthonormal basis for the subspace spanned by them. The set \( \{2\mathbf{e}_1, 2\mathbf{e}_2, \ldots, 2\mathbf{e}_m\} \subseteq \mathbb{R}^m \) is an orthogonal, but not orthonormal basis for \( \mathbb{R}^m \).
• Determine whether the following set is orthogonal and/or orthonormal:

\[
\{[2, 0, -1]^T, [1, 1, -1]^T, [-1, -1, -2]^T\}
\]

• The following is a basis for the subspace \( W \subseteq \mathbb{R}^3 \) of all vectors having 0 for their third entry:

\[
\{[1, -2, 0]^T, [-2, -1, 0]^T\}
\]

Determine whether it is an orthogonal and/or orthonormal basis.
If we have an orthogonal basis $B$ for some subspace $W$ and also a vector $v \in W$, the unique coefficients that are used to express $v$ as a linear combination of the vectors in $B$ are easily computed. This is a consequence of the following theorem, that works even if $B$ is not a basis:

**THEOREM (coefficients with respect to an orthogonal set)**

If a vector $v$ is a linear combination of an orthogonal set of nonzero vectors $u_1, u_2, \ldots, u_k$, then this combination is given by

$$v = \left( \frac{v \cdot u_1}{\|u_1\|^2} \right) u_1 + \left( \frac{v \cdot u_2}{\|u_2\|^2} \right) u_2 + \cdots + \left( \frac{v \cdot u_k}{\|u_k\|^2} \right) u_k$$

Proof. Let $v = c_1 u_1 + c_2 u_2 + \cdots + c_k u_k$. Taking the dot product with $u_1$, only the first term survives due to orthogonality of the $u_i$'s:

$$v \cdot u_1 = c_1 u_1 \cdot u_1 + c_2 u_2 \cdot u_1 + \cdots + c_k u_k \cdot u_1 = c_1 \|u_1\|^2 + c_2 0 + \cdots + c_k 0 = c_1 \|u_1\|^2$$

Since $u_1 \neq 0$, its norm is not zero and we can divide by it:

$$c_1 = \frac{v \cdot u_1}{\|u_1\|^2}$$

Take dot products with $u_2, \ldots, u_k$ to obtain $c_2, \ldots, c_k$ similarly.

**PROBLEM.** First verify that $v = [5, -1, 0]^T$ belongs to the subspace

$$W = \text{span}\{[1, -2, 3]^T, [3, 0, -1]^T\}$$

Then verify that the vectors above form an orthogonal basis for $W$, and use the formula above to find the coefficients of $v$ as a linear combination of them.

If we have an orthogonal set not containing $0$, then each vector can be pictured as an arrow that is perpendicular to all the other arrows in this set. It should be clear that, because of this, none of these arrows is a linear combination of the others, so the set must be independent. In the following theorem we prove this analytically:

**THEOREM (linear independence of nonzero orthogonal vectors)**

Any orthogonal set not containing $0$ is linearly independent.

Proof. We must show that, if scalars $c_1, c_2, \ldots, c_k$ are such that

$$0 = c_1 u_1 + c_2 u_2 + \cdots + c_k u_k$$

then they must be all zero. But the previous theorem tells us that there is a formula for these scalars:

$$c_i = \frac{0 \cdot u_i}{\|u_i\|^2} = 0$$

as we wanted.

Being able to easily find the coefficients in a linear combination is just one of the many uses for having an orthogonal or orthonormal basis for a subspace. So it’s good to know that we can always find such types of bases:

**ALGORITHM (Gram-Schmidt)**

See theorem 6.6 on page 378 for a proof. The tricky part is showing that each step preserves the span of the subset of vectors considered thus far.

Every subspace $W \subseteq \mathbb{R}^m$ admits an orthogonal (and an orthonormal) basis. To find one:

- Start with any basis $\{u_1, u_2, \ldots, u_k\}$.
- Define $v_1 = u_1$.
- Define $v_2 = u_2 - \left( \frac{u_2 \cdot v_1}{\|v_1\|^2} \right) v_1$.
- Define $v_3 = u_3 - \left( \frac{u_3 \cdot v_1}{\|v_1\|^2} \right) v_1 - \left( \frac{u_3 \cdot v_2}{\|v_2\|^2} \right) v_2$. 


• Proceed until the last vector: $v_k = u_k - \left( \frac{u_k \cdot v_1}{\|v_1\|^2} \right) v_1 - \left( \frac{u_k \cdot v_2}{\|v_2\|^2} \right) v_2 - \cdots - \left( \frac{u_k \cdot v_{k-1}}{\|v_{k-1}\|^2} \right) v_{k-1}$.

• The set $\{v_1, v_2, \ldots, v_k\}$ is an orthogonal basis for $W$.

• If we need an orthonormal basis, normalize each vector: $\left\{ \frac{v_1}{\|v_1\|}, \frac{v_2}{\|v_2\|}, \ldots, \frac{v_k}{\|v_k\|} \right\}$.

As you calculate each $v_i$, it’s always a good idea to take the dot product with all previous $v_j$’s to see that they are indeed orthogonal.

**PROBLEM.** Find an orthonormal basis for

$$\text{span} \left\{ \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -3 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \\ 3 \\ -4 \end{bmatrix} \right\}$$

Then use the formula introduced in this section to express $[0, 9, 0, 1]^T$ as a linear combination of the vectors in that orthonormal basis (you are told that this vector belongs to $W$, no need to check).

**Suggested problems from the book**

- Systematic: 1-24
- True or false: 41-51
- Conceptual: 53

---

### 6.3. Orthogonal projections

**Key concepts:** orthogonal projection onto a subspace, projection matrix, orthogonal complement, distance to a subspace

We learned how to compute the orthogonal projection of a point onto a line going through the origin. Such a line is a subspace, and it also makes sense to want to project a point onto a more general subspace, of higher dimension, like a plane inside 3D space. Thinking geometrically, we can probably guess that the orthogonal projection of $P$ onto a subspace $W$ must be the point $P_0$ in $W$ closest to $P$. But let’s simply give an analytical definition and just remark later that it indeed satisfies this property:

**DEFINITION** (orthogonal projection onto a subspace)

Let $W \subseteq \mathbb{R}^m$ be a subspace different from $\{0\}$, let $u \in \mathbb{R}^m$ be any vector. The orthogonal projection or simply projection of $u$ onto $W$ is the vector $\mathbb{P}_W(u) \in W$ defined as

$$\mathbb{P}_W(u) = \left( \frac{u \cdot w_1}{\|w_1\|^2} \right) w_1 + \left( \frac{u \cdot w_2}{\|w_2\|^2} \right) w_2 + \cdots + \left( \frac{u \cdot w_k}{\|w_k\|^2} \right) w_k$$

where $\{w_1, w_2, \ldots, w_k\}$ is any orthogonal basis for $W$. It can be shown that this is independent of the particular orthogonal basis chosen for $W$, but it must be orthogonal!

(The book calls this $\mathbb{P}_W(u)$, not a fancy symbol $\mathbb{P}_W(u)$).

- Find (and draw) the projection of $(1, 2, 3)$ onto the xy-plane, that is, the subspace spanned by $[1, 0, 0]^T$ and $[0, 1, 0]^T$.
- Compute the projection of $[1, 2, 3]^T$ onto the subspace $W = \text{span}\{[3, -3, 0]^T, [0, 1, -1]^T\}$

Note that these generating vectors are not orthogonal, so you’ll first need to find an orthogonal basis.
ORTHOGONALITY

It can be easily checked that the formula for the function \( P_W \) satisfies the linearity properties, therefore it is a linear transformation, and must have an associated standard matrix.

Let \( \{ w_1, w_2, \ldots, w_k \} \) be a basis (doesn’t even need to be orthogonal) for \( W \). Let \( C \) denote the matrix having them as columns:

\[
C = \begin{bmatrix} w_1 & w_2 & \cdots & w_k \end{bmatrix}
\]

Then the standard matrix of the linear transformation \( P_W \) is

\[
C(C^T C)^{-1} C^T
\]

(\text{Note that } C^T C \text{ is square, even if } C \text{ and } C^T \text{ are not. It can be proven (see lemma on page 395) that it’s invertible when the columns of } C \text{ are independent, as they are in this application.)}

Commit this to memory: you write \( C^T C \) first, then you take the inverse of the product of the two matrices in the middle. A proof of this formula is at the end of this section.

- If we already have an orthogonal basis for \( W \) and want to compute the projection of some vector onto \( W \), it’s usually easier to apply the formula

\[
P_W(u) = \left( \frac{u \cdot w_1}{\|w_1\|^2} w_1 + \frac{u \cdot w_2}{\|w_2\|^2} w_2 + \cdots + \frac{u \cdot w_k}{\|w_k\|^2} w_k \right)
\]

If we just have a basis, we can do Gram-Schmidt to make it orthogonal and apply this formula, or we can use the vectors in the basis to form the \( C \) matrix and apply the formula \( P_W(u) = (C^T C)^{-1} C^T u \). Apply this to the last example.

- Find the orthogonal projection matrix \( P_W \) for the subspace \( \text{Col} \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 2 & 5 \end{bmatrix} \). Use it to find the distance between \( W \) and the point \((-6, 4, 5)\).

\[\text{Can’t simplify the formula fot the matrix of a projection!}\]

We’ve learned that the inverse of a product is the product of the inverses in the reverse order, so we might think about simplifying the formula above as

\[
C(C^T C)^{-1} C^T = C(C^{-1} (C^T)^{-1}) C^T = (CC^{-1}) ((C^T)^{-1} C^T) = II = I
\]

But this is not valid! The property

\[
(AB)^{-1} = B^{-1} A^{-1}
\]

is only true for square, invertible matrices of the same size. In our case, \( C \) is formed by elements of a basis for \( W \subseteq \mathbb{R}^m \) as columns, so it is a \( m \times (\dim W) \) matrix, and in general \( W \) is not the entire \( \mathbb{R}^m \), so \( \dim W < m \). That is, our \( C \) and \( C^T \) are not square.

\[\text{Projection matrix onto a 1D subspace}\]

If \( W = \text{span}\{u\} \), the standard matrix of \( P_W \) is given by

\[
A = \frac{1}{\|u\|^2} uu^T
\]

(Notice that the matrix product \( uu^T \), of an \( m \times 1 \) matrix by a \( 1 \times m \) matrix, is well defined and yields an \( m \times m \) matrix). This matrix satisfies \( A^2 = A \) and \( A^T = A \).

But this topic was already covered in the last section, when we talked about projecting onto a line through the origin. This formula is never used in practice, only to prove theorems.

\[\text{Justification.}\] The matrix \( C \) contains only one column with \( u \) in it, that is, \( C = u \). Since \( u^T u = u \cdot u = \|u\|^2 \), the projection matrix becomes

\[
u(u^T u)^{-1} u^T = u \frac{1}{\|u\|^2} u^T = \frac{1}{\|u\|^2} uu^T
\]

Use the fact that \( \|u\|^2 = u^T u \) to show that \( A^2 = A \), and the property \( (XY)^T = Y^T X^T \) to show that \( A^T = A \).
DEFINITION (orthogonal complement)  
Let \( S \subseteq \mathbb{R}^m \) be a nonempty set. The **orthogonal complement** of \( S \), denoted \( S^\perp \) and pronounced “\( S \) perp”, is the subspace (we must show that it is always a subspace) consisting of all vectors that are orthogonal to every vector in \( S \):

\[
S^\perp = \{ u \in \mathbb{R}^m \mid u \cdot v = 0 \text{ for all } v \in S \}
\]

If \( S \) is a span:

\[
S = \text{span}\{v_1, v_2, \ldots, v_k\}
\]

then, by linearity properties of the dot product, a vector is orthogonal to everything in \( S \) if and only if it is orthogonal to each vector in the generating set, so we can also write

\[
S^\perp = \{ u \in \mathbb{R}^m \mid u \cdot v_i = 0 \text{ for each } i = 1, 2, \ldots, k \}
\]

This gives us a way to calculate \( S^\perp \): each condition \( u \cdot v_i = 0 \) is a linear equation in the entries \( u_1, u_2, \ldots, u_m \). All together they form a (homogeneous) linear system that we can solve to find \( S^\perp \).

- In the space \( \mathbb{R}^m \), we have \( \{0\}^\perp = \mathbb{R}^m \) (the set of the vectors orthogonal to \( 0 \) is everything) and \( (\mathbb{R}^m)^\perp = \{0\} \) (the set of vectors orthogonal to all vectors only contains 0).
- In the space \( \mathbb{R}^3 \), if \( W \) denotes the \( xy \)-plane (\( W = \text{span}\{(1,0,0)^T, (0,1,0)^T\} \)), then \( W^\perp \) is the \( z \)-axis (\( W^\perp = \text{span}\{(0,0,1)^T\} \)). Indeed, the only vectors orthogonal to any vector on the \( xy \)-plane are the ones along the \( z \)-axis. (Picture!)

 Decomposition of a vector into a part in a subspace and a part orthogonal to it  
Given a vector \( u \in \mathbb{R}^m \) and a subspace \( W \subseteq \mathbb{R}^m \), there is a unique way to write \( u \) as a sum of a vector \( u_1 \) in \( W \) and a vector \( u_2 \) in \( W^\perp \):

\[
\begin{align*}
    u_1 &= \mathbb{P}_W(u) \\
    u_2 &= u - \mathbb{P}_W(u)
\end{align*}
\]

Let \( P \) and \( P_0 \) be the points associated to vectors \( u \) and \( \mathbb{P}_W(u) \). Then \( P_0 \) is the closest point to \( P \) in \( W \), and the distance from \( P \) to \( W \) is therefore defined as its distance to \( P_0 \):

\[
\text{dist}(P, W) = |u - \mathbb{P}_W(u)|
\]

**PROBLEM.** Find the distance between the point \((-3, 2, -4, -1)\) and the subspace

\[
W = \text{span}\{(1, -1, 1, -1)^T, [-1, -1, 1, 1]^T\}
\]

Now for some theorems:

**THEOREM**

a) If \( W \subseteq \mathbb{R}^m \) is a subspace, then \( (W^\perp)^\perp = W \). This is not true for a set \( W \) that is not a subspace.

b) For any matrix \( A \), we have \( \text{Row}(A) = (\text{Null}(A))^\perp \).

c) Let \( W \subseteq \mathbb{R}^m \) be a subspace. Then \( \dim(W) + \dim(W^\perp) = m \).

**Proof.**

- a) By definition, \( (W^\perp)^\perp \) is the set of all vectors orthogonal to all vectors orthogonal to all vectors in \( W \). It is a logical consequence (may make your head hurt a little) that the vectors in \( W \) satisfy this property, so that

\[
W \subseteq (W^\perp)^\perp
\]
The proof that no other vectors satisfy this property is a little more subtle (this is where we need to use the fact that \( W \) is a subspace) and we won’t include it here.

b) \( \text{Row}(A) \) is spanned by the rows of \( A \), so \( (\text{Row}(A))^{\perp} \) can be found by solving the system \( Ax = 0 \). But the solutions of this are precisely the elements of \( \text{Null}(A) \). It follows that
\[
(\text{Row}(A))^{\perp} = \text{Null}(A)
\]
Now take \( \perp \) on both sides (use item (a)).

c) Take a basis for \( W \) and form a matrix \( A \) with its vectors as rows. Then \( \dim(W) = \dim(\text{Row}(A)) = \dim(\text{Col}(A)) = \text{rank}(A) \). By the previous theorem, \( W^{\perp} = \text{Null}(A) \), so \( \dim(W^{\perp}) = \text{nullity}(A) \). And the sum of rank and nullity is the number of columns of \( A \), that is, the number of entries in each element of \( W \) (since the elements of a basis for \( W \) were put into the rows of \( A \)), which is \( m \).

Proof of the formula for the matrix of a projection. Fix a vector \( u \in \mathbb{R}^m \) to be projected onto \( W \). Since \( P_W(u) \) is an element of \( W \), it is in the span of a given basis for \( W \), that is, the system
\[
Cx = P_W(u)
\]
(where \( C \) has the basis elements as columns) has a solution \( x \). So we can write \( Cx \) instead of \( P_W(u) \).

The defining property for \( P_W(u) \) is that \( u - P_W(u) \in W^{\perp} \). Now \( W^{\perp} \) is the same as \( (\text{Col}(C))^{\perp} = (\text{Row}(C^T))^{\perp} = \text{Null}(C^T) \). That means \( C^T \) applied to \( u - P_W(u) = u - Cx \) must be 0:
\[
C^T(u - Cx) = 0 \implies C^T u = C^T C x
\]
Turns out \( C^T C \) is invertible, as remarked above, so we can solve for \( x \):
\[
x = (C^T C)^{-1} C^T u
\]
But then \( P_W(u) = Cx = C(C^T C)^{-1} C^T u \). So we have found the matrix \( A \) such that \( P_W(u) = Au \) for any \( u \), and it is
\[
A = C(C^T C)^{-1} C^T
\]

Suggested problems from the book
- Systematic: 1-32
- True or false: 35-37, 39-49, 52-56
- Conceptual: 61-62, 65, 67

6.4. Least Squares Algorithm

Key concepts: Least Squares Algorithm for a line and for a general curve

MOTIVATION

Suppose we want to find the coefficients \( A, B \) of the line
\[
L : y = Ax + B
\]
that, in some sense, most closely approximates the data points
\[
\{(1,1), (2, 4), (3, 9)\}
\]
They clearly belong to a parabola, but no line.

If \( (x_0, y_0) \) denotes any one of these points, then the value
\[
y_0 - (Ax_0 + B)
\]
is zero if and only if \((x_0, y_0) \in L\). If this value is not zero, but small, then the point \((x_0, y_0)\) is not on \(L\) but is close to it. So we can think of this value as a measure of the error committed when using line \(L\) to describe the relationship between \(x_0\) and \(y_0\).

At first, then, it might seem that what we need is to find \(A, B\) such that the sum of the corresponding errors for each point is the least possible:

\[
\left(1 - (A + B)\right) + \left(4 - (2A + B)\right) + \left(9 - (3A + B)\right)
\]

But this wouldn’t be quite right. This sum might be very small even if some of the individual errors are large, because they can have different signs and cancel out. So instead of minimizing the sum of these values, we try to minimize the sum of their squares (these are always \(\geq 0\), so there’s no cancellation). The problem becomes: find \(A, B\) such that the sum

\[
E = \left(1 - (A + B)\right)^2 + \left(4 - (2A + B)\right)^2 + \left(9 - (3A + B)\right)^2
\]

is the least possible. Now this expression is just the distance squared between the 3D vectors

\[
[1, 4, 9]^T \quad \text{and} \quad [A + B, 2A + B, 3A + B]^T
\]

The set of all possible values for the second vector, as \(A, B\) vary, is spanned by the vectors \([1, 1, 1]^T\) and \([1, 2, 3]^T\), because

\[
\begin{bmatrix}
A + B \\
2A + B \\
3A + B
\end{bmatrix}
= B
\begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix}
+ A
\begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix}
\]

And we just learned that the vector in

\[W = \text{span}\{[1, 1, 1]^T, [1, 2, 3]^T\}\]

that minimizes \(E\) is the projection \(P_0\) of \(P = (1, 4, 9)\) onto \(W\) (we are trying to minimize the distance squared, but of course the point that realizes the minimum distance squared will also realize the minimum distance). We also know how to compute it using the projection matrix:

\[
P_0 = P(W) = C(C^T C)^{-1} C^T [1, 4, 9]^T
\]

where \(C\) has a basis for \(W\) as its columns:

\[
C = \begin{bmatrix}
1 & 1 \\
1 & 2 \\
1 & 3
\end{bmatrix}
\]

But \(P_0\) is not what we want. \(P_0\) is the vector \([A + B, 2A + B, 3A + B]^T\), and we actually just want its coefficients \(A, B\). Well, the vector \([P_0]_B\) containing its coefficients in the basis \(B = \{[1, 1, 1]^T, [1, 2, 3]^T\}\) is related to \(P_0\) by

\[
P_0 = C[P_0]_B
\]

Therefore, multiplying by \(C^T\) on both sides:

\[
C^T C[P_0]_B = C^T P_0 = C^T C(C^T C)^{-1} C^T [1, 4, 9]^T = C^T [1, 4, 9]^T
\]

Finally, we remarked in the last section that \(C^T C\) is always invertible when the columns of \(C\) are independent. So we can move it to the right-hand side as an inverse:

\[
[P_0]_B = (C^T C)^{-1} C^T [1, 4, 9]^T
\]

That is,

\[
\begin{bmatrix}
A \\
B
\end{bmatrix} = \left(\begin{bmatrix}
1 & 1 & 1 \\
1 & 2 & 1 \\
1 & 3
\end{bmatrix}\right)^{-1} \begin{bmatrix}
1 & 1 & 1 \\
1 & 2 & 1 \\
1 & 3
\end{bmatrix} \begin{bmatrix}
1 \\
4 \\
9
\end{bmatrix} = \cdots
\]
The same reasoning as above can be used to deduce the following:

**ALGORITHM (Least Squares)**

Suppose we want to approximate a collection of \( m \) data points

\[
\{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\} \quad (\text{all } x_i \text{ are distinct})
\]

with a curve of degree \( n < m - 1 \):

\[
y = P(x) = A_0 + A_1 x + A_2 x^2 + \cdots + A_n x^n
\]

We want to find the coefficients \( A_i \) of the curve that minimizes the **error sum of squares**

\[
E = (y_1 - P(x_1))^2 + (y_2 - P(x_2))^2 + \cdots + (y_m - P(x_m))^2
\]

The curve is called the **least-squares fit** of degree \( n \).

Construct a matrix \( C_{m \times (n+1)} \) containing the powers of the \( x_i \)'s as its columns:

\[
C = \begin{bmatrix}
1 & x_1 & x_1^2 & \cdots & x_1^n \\
1 & x_2 & x_2^2 & \cdots & x_2^n \\
1 & x_3 & x_3^2 & \cdots & x_3^n \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_m & x_m^2 & \cdots & x_m^n 
\end{bmatrix}
\]

Then the coefficients \( A_i \) are given by

\[
\begin{bmatrix}
A_0 \\
A_1 \\
A_2 \\
\vdots \\
A_n
\end{bmatrix} = (C^T C)^{-1} C^T \begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
\vdots \\
y_m
\end{bmatrix}
\]

**Note:** Nothing in this uses \( n < m - 1 \). We usually consider this condition because, if \( n \) were \( m - 1 \) or larger, then \( n + 1 \) would be \( m \) or larger, that is, there would be at least as many degrees of freedom in the curve as data points (a curve of degree \( n \) has \( n + 1 \) free coefficients), so we would actually be able to find an exact fit. In statistical problems we usually have many more points to approximate than degrees of freedom in the curve.

- Find the best-fit lines for each of the following collections of data points:
  
  \( (1, 2), (2, 4), (3, 5) \)
  
  \( (1, 2), (2, 4), (3, 6), (5, 10) \)

- Find the best-fit parabola: \((-2, 0), (0, 5), (1, 5), (3, 12)\)

**Remark** (curves other than polynomials)

This method can also be used to approximate a collection of points by more general curves (under certain technical assumptions). For example, to approximate a collection by a curve of the form \( F(x) = A \log x + Be^x + C \sin(x) \), you would use the matrix

\[
C = \begin{bmatrix}
1 & \log x_1 & e^{x_1} & \sin(x_1) \\
1 & \log x_2 & e^{x_2} & \sin(x_2) \\
\vdots & \vdots & \vdots & \vdots \\
1 & \log x_m & e^{x_m} & \sin(x_m)
\end{bmatrix}
\]

But this will not be covered in this class.

**Suggested problems from the book**

- **Systematic:** 1-8, 10-15
DEFINITION (orthogonal matrix)

Orthogonal matrix: A square, invertible matrix $A$ such that $A^{-1} = A^T$. Equivalently, an $m \times m$ matrix whose columns form an orthonormal set. Equivalently, an $m \times m$ matrix whose rows form an orthonormal set.

Careful! The matrix is called orthogonal, even though its columns and rows are orthonormal!

The usual symbol for an orthogonal matrix is $O$, but we have been using that to denote the zero matrix, so we’ll use $A$ here.

- $I_m$ is orthogonal, as is any matrix where each row and column has only one 1 and everything else 0, for example

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

It’s easy to see that such matrices are orthogonal by checking that the rows (or columns) are orthonormal. But it’s not so easy to check that $A^{-1} = A^T$ (it is true, though).

- The rotation matrices $R_\theta$ in $\mathbb{R}^2$ are orthogonal.

- A reflection matrix in $\mathbb{R}^2$, associated to the operation of reflecting about a line through the origin, is orthogonal. To see this, first note that the matrix that reflects about the $x$-axis is orthogonal:

$$B = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

A general reflection $A$ about some other line can be obtained from $B$ by a rotation change of coordinates; that is, $A$ and $B$ are similar through a change-of-coordinates matrix $R_\theta$:

$$A = R_\theta^{-1}BR_\theta$$

We can use this to see that $A^{-1} = A^T$, that is, to check that $AA^T = I$:

$$AA^T = (R_\theta^{-1}BR_\theta)(R_\theta^{-1}BR_\theta)^T = (R_\theta^{-1}BR_\theta)(R_\theta^T B^T (R_\theta^{-1})^T)$$

The product $R_\theta R_\theta^T$ inside cancels out (is equal to $I$) because $R_\theta$ is orthogonal. Then the $BB^T$ product also drops out, because $B$ is orthogonal. Finally, $R_\theta^{-1}(R_\theta^{-1})^T$ is also $I$, because $R_\theta^{-1}$ is also orthogonal (it is also a rotation matrix).

- Here’s a justification for why the property $A^{-1} = A^T$ (which is the same as $AA^T = I$) is equivalent to saying that the rows of $A$ are an orthonormal set. Take a $3 \times 3$ matrix, for example:

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Suppose that $AA^T = I$:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

If we compute any element of a position $(i, i)$ on the left side, for example the $(2, 2)$ element $d^2 + e^2 + f^2$, the equation is telling us that it is equal to 1 on the right. But we recognize $d^2 + e^2 + f^2$ as the norm-square of the second row. Therefore all rows have norm-square 1, and hence also norm 1.

If instead we compute any element of a position $(i, j)$ on the left side with $i \neq j$, for example the $(1, 3)$ element $ag + bh + ci$, the equation is telling us that it is equal to 0 on the right. But we recognize $ag + bh + ci$ as the dot product between the first and third rows. Therefore different rows have dot product 0, and hence are orthogonal to each other. That is, the set of the rows is orthonormal. Use now the product $A^T A = I$ to conclude that also the set of the columns is orthonormal.

6.5. Orthogonal matrices

Key concepts: orthogonal matrices, preservation of dot product and norm, orthogonal matrices in the plane
Orthogonal matrices must have determinant \( \pm 1 \) and all entries between \(-1\) and \(1\) (inclusive). But not all matrices with either or both of these properties are orthogonal.

**Justification.** Since \( A^T = A^{-1} \), we have \( AA^T = I \). Taking determinants, \( \det(A) \det(A^T) = \det(I) = 1 \). Now use \( \det(A^T) = \det(A) \) to transform this into \( \det(A)^2 = 1 \), whose solutions are \( \det(A) = \pm 1 \).

The sum of the squares of the entries in any given row or column of \( A \) is \( 1 \); they must all be numbers in the range \(-1 \leq x \leq 1\).

**PROBLEM.** Find an orthogonal matrix \( A_{3\times 3} \) such that

\[
A = \begin{bmatrix}
0.6 & 0 \\
-0.8 & 0 \\
0 & 1
\end{bmatrix}
\]

(There are many possible answers.)

**SOLUTION.** The trick is to realize that the condition required is the same as saying that the third row of \( A \) is \([0.6, -0.8, 0]^T\). Indeed, because \( A[0.6, -0.8, 0]^T = [0, 0, 1]^T = e_3 \), and because we also want \( A^T A = I \) (so that \( A \) is orthogonal), we must have

The 3rd row of \( A \) = 3rd column of \( A^T e_3 = A^T A[0.6, -0.8, 0]^T = I[0.6, -0.8, 0]^T = [0.6, -0.8, 0]^T \)

Then we just have to select the other two rows of \( A^T \) in a way that all 3 rows form an orthonormal set. So they will need to be an orthonormal basis for the complement \( \{[0.6, -0.8, 0]^T\}^\perp \).

The following result is the reason why orthogonal matrices are important:

**THEOREM (orthogonal matrices preserve the Euclidean geometry)**

Let \( A_{m \times m} \) be an orthogonal matrix. Then the following are true:

- As a linear transformation, \( A \) preserves the inner product of any two vectors:
  
  For any \( u, v \in \mathbb{R}^m \), \( \langle A u, A v \rangle = \langle u, v \rangle \)

- As a linear transformation, \( A \) preserves the norm of any vector:
  
  For any \( u \in \mathbb{R}^m \), \( \| A u \| = \| u \| \)

This theorem actually goes both ways: any matrix satisfying one of these properties must be orthogonal.

**Proof.** Suppose \( A \) is orthogonal: \( A^T A = AA^T = I \). Use the property that a matrix can be moved to the other side of an inner product as a transpose:

\[
(A u) \cdot (A v) = u \cdot (A^T A v) = u \cdot (I v) = u \cdot v
\]

This proves the first item. The second is a consequence of the first together with the property

\[
\| u \|^2 = u \cdot u
\]

The proof that the theorem goes both ways involves these same properties, plus some understanding of how to use inner products to conclude that two matrices are equal. Let’s not worry about it.

The examples we gave of orthogonal \( 2 \times 2 \) matrices included rotations and reflections. It turns out that those are the only possible types:

**THEOREM (2D orthogonal matrices)**

a) Let \( A_{2 \times 2} \) be an orthogonal matrix. Then:

- If \( \det A = 1 \), then \( A \) is a rotation matrix.
- If \( \det A = -1 \), then \( A \) is a reflection matrix.

b) Let \( A, B \) be \( 2 \times 2 \) orthogonal matrices. Then:
– If both are reflections, then \( AB \) is a rotation.
– If one is a reflection and the other a rotation, then \( AB \) is a reflection.

Proof of (b). First show that \( AB \) is also orthogonal, then use (a) together with the property

\[
\det(AB) = \det(A) \det(B)
\]

The proof of (a) is actually quite technical and not very illuminating (see Theorem 6.11 in the book on page 416). But it’s at least intuitive, because we just learned that orthogonal transformations preserve norms and dot products, therefore also lengths and angles. Try to imagine a transformation of the plane that first of all takes \( 0 \) to \( 0 \) (it must be linear) and also doesn’t bring apart any two arbitrary points (because of this property it is called a rigid motion), so that all angles, lengths and areas are preserved. It could be a rotation of the whole plane or a reflection, but nothing else.

**PROBLEM.** Each of the matrices below is orthogonal. Determine if they are a rotation (in that case, determine also the angle) or a reflection (in that case, determine also the line of reflection).

\[
A = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix} \quad B = \begin{bmatrix} -0.6 & 0.8 \\ -0.8 & -0.6 \end{bmatrix}
\]

**SOLUTION.** Take determinants to see which is which. The angle of a rotation matrix is easy to find, because we know the format of a rotation matrix, so we can find \( \cos \theta \) and \( \sin \theta \), and from there also \( \theta \). To find the line of reflection of a reflection matrix, notice that, if \( X \) is a reflection matrix about a line spanned by a vector \( u \), then \( Xu = u \) (that is, \( u \) reflected about its line is itself), which means that \( u \) is an e-vector corresponding to e-value 1.

**REMARK (orthogonal matrices in higher dimensions)**
Orthogonal matrices of higher dimensions are not so easy to describe as in the \( 2 \times 2 \) case. The so-called orthogonal group \( O(n) \) (the set of all orthogonal \( n \times n \) matrices) is an active field of research with applications even to quantum physics.

**Orthogonal operators**
As you can probably guess, an orthogonal operator \( T : \mathbb{R}^m \to \mathbb{R}^m \) is an operator whose standard matrix is orthogonal. Problems in this section can be phrased about matrices or operators, but the two cases are completely analogous.

**Suggested problems from the book**
- Systematic: 1-16, 37-38, 42
- True or false: 17-31
- Conceptual: 39, 43, 51

6.6. Symmetric matrices

Key concepts: symmetric matrices, diagonalization of symmetric matrices, spectral decomposition, rotation of ellipses

**DEFINITION (symmetric matrix)**
Symmetric matrix: A square matrix \( A \) such that \( A = A^T \).
Therefore, \( a_{ij} = a_{ji} \), that is, the entries of \( A \) are symmetric with respect to the main diagonal.
• A $5 \times 5$ symmetric matrix:

\[
\begin{bmatrix}
4 & 0 & -2 & 8 & -9 \\
0 & 2 & -2 & 5 \\
-2 & 2 & 0 & 1 & 6 \\
8 & -2 & 1 & -1 & -7 \\
-9 & 5 & 6 & -7 & e
\end{bmatrix}
\]

• Any diagonal matrix is symmetric.

• Prove that, if $A, B$ are symmetric, then $AB + BA$ is also symmetric. Solution:

$$(AB + BA)^T = (AB)^T + (BA)^T = B^TA^T + A^TB^T = BA + AB = AB + BA$$

(be careful with the order!)

• Find a diagonalization of the following symmetric matrix:

\[
\begin{bmatrix}
2 & -2 \\
-2 & 5
\end{bmatrix}
\]

What can you say about the matrix $P$?

The relation between symmetric matrices and orthogonality, hinted at in the last example above, is given by the following two theorems:

**THEOREM (orthogonality of e-vectors for symmetric matrices)**

If $A_{m \times m}$ is symmetric, then e-vectors corresponding to different e-values are orthogonal.

*Proof.* Suppose $A = A^T$, $Au = \lambda_1 u$ and $Av = \lambda_2 v$, with $\lambda_1 \neq \lambda_2$. That is, $A$ is symmetric and $u, v$ are e-vectors corresponding to distinct e-values. We must show that $u \cdot v = 0$.

First simplify $(Au) \cdot v$ using $Au = \lambda_1 u$. Next, move the $A$ to the other side of the inner product:

$$(Au) \cdot v = u \cdot (A^Tv) = u \cdot (Av)$$

and simplify this using $Av = \lambda_2 v$. You’ll reach the conclusion that

$$\lambda_1 (u \cdot v) = \lambda_2 (u \cdot v)$$

If $u \cdot v$ were not 0, then we could cancel it from this equation to obtain the contradiction $\lambda_1 = \lambda_2$. Therefore it must be that $u \cdot v = 0$.

**THEOREM (diagonalization with orthogonal e-vectors)**

Let $A_{m \times m}$ be a square matrix. Then $A$ is symmetric if and only if there exists an orthonormal basis for $\mathbb{R}^m$ consisting of e-vectors of $A$.

In particular, every symmetric matrix is diagonalizable, and a diagonalization for it can be written in the form $PDP^T$ instead of $PDP^{-1}$, because $P$ can be chosen as an orthogonal matrix ($P^{-1} = P^T$) if we select e-vectors of norm 1.

*Proof.* One side is easy: if there exists an orthonormal basis of e-vectors of some matrix $A$, then $A$ is diagonalizable and the matrix $P$ in its diagonalization is orthogonal (its columns form an orthonormal set). Then

$$A = PDP^{-1} = PDP^T$$

If we try to find $A^T$, we get

$$A^T = (PDP^T)^T = (P^T)^TDP^T = PDP^T = A$$

Therefore $A$ is symmetric.

The converse statement is a much deeper result whose proof we omit.

With this theorem, we can prove that every symmetric matrix is the sum of projection matrices multiplied by scalars in a special way. This is the deepest result in this course, but we won’t have any application for it (it is used to prove more advanced theorems about symmetric matrices).
THEOREM (spectral theorem for symmetric matrices)

Every symmetric matrix $A_{m \times m}$ can be written as a sum of at most $m$ terms

$$A = c_1 P_1 + c_2 P_2 + \cdots + c_k P_k$$

called a spectral decomposition of $A$, where:

- Each $P_i$ is a rank 1 projection matrix (that is, $P_i^2 = P_i$ and $P_i^T = P_i$), each $c_i$ is a scalar.
- The projections are onto different lines, that is, $P_i P_j = O$ if $i \neq j$.

To obtain a spectral decomposition, take the $c_i$’s to be the e-values of $A$, and each $P_i$ to be the projection matrix onto a line spanned by a corresponding e-vector $v_i$, which is $P_i = v_i v_i^T$ if it is a unit vector.

Proof. Let $A = PDP^T$ where $D$ is diagonal with the e-values $\lambda_1, \lambda_2, \ldots, \lambda_m$ along the main diagonal, and $P$ is orthogonal with a basis of corresponding e-vectors $v_1, v_2, \ldots, v_m$ as its columns. Then

$$A = PDP^T$$

$$= \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_m^T \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 v_1^T \\ \lambda_2 v_2^T \\ \vdots \\ \lambda_m v_m^T \end{bmatrix}$$

$$= \begin{bmatrix} v_1 \lambda_1 v_1^T + v_2 \lambda_2 v_2^T + \cdots + v_m \lambda_m v_m^T \end{bmatrix}$$

$$= \lambda_1 [v_1 v_1^T] + \lambda_2 [v_2 v_2^T] + \cdots + \lambda_m [v_m v_m^T]$$

The matrices $P_j = v_j v_j^T$ are the projection matrices onto the 1D subspaces spanned by each $v_j$, according to the formula we pointed out earlier in this chapter (there is no need to divide by $\|v_i\|^2$ since the $v_i$’s are unit vectors). If we try to multiply together any two of them, we get $O$ because $v_i \cdot v_j = 0$ when $i \neq j$:

$$P_i P_j = (v_i v_i^T)(v_j v_j^T) = v_i (v_i^T v_j) v_j^T = v_i v_i v_j v_j = 0$$

- Find a spectral decomposition of the matrix

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

The e-values and corresponding e-vectors can be taken as

$$\lambda_1 = 3 \quad , \quad \lambda_2 = 3 \quad , \quad \lambda_3 = 6$$

$$u_1 = [1, 0, 1]^T \quad , \quad u_2 = [-1, 2, 1]^T \quad , \quad u_3 = [1, 1, -1]^T$$

- Show that

$$B = \begin{bmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{bmatrix} - 5 \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix} = \begin{bmatrix} 3 & -4 \\ -4 & -3 \end{bmatrix}$$

is a spectral decomposition of the matrix
Now we go back to the topic of rotation of ellipses. In chapter 4 we learned how to find the complicated equation of an ellipse in a rotated reference frame. Now, what we want to be able to do is the converse: given the complicated equation of a rotated ellipse, find the angle of rotation that turns it into a standard ellipse.

**Algorithm (rotating ellipse into a standard ellipse)**

Given the equation of a rotated ellipse

\[ Ax^2 + Bxy + Cy^2 = 1 \]

consider \( a = A, b = B/2, c = C \) and the symmetric matrix

\[
\begin{bmatrix}
    a & b \\
    b & c
\end{bmatrix}
\]

Find a diagonalization \( PDP^T \) of it. Then the equation of the ellipse in the coordinates \((u,v)\) defined by

\[
\begin{bmatrix}
    u \\
    v
\end{bmatrix} = P^T \begin{bmatrix} x \\ y \end{bmatrix}
\]

will be

\[ \lambda_1 u^2 + \lambda_2 v^2 = 1 \]

where \( \lambda_1, \lambda_2 \) are the entries along the diagonal of \( D \). Furthermore, if \( \det(P) = 1 \) (if you find \( \det(P) = -1 \), change the order of the e-values in \( D \) and the columns in \( P \)), then \( P \) is a rotation matrix \( R_\theta \), and the angle that the ellipse has been rotated counterclockwise from a standard position is \(-\theta\).

- Draw the ellipse \( 2x^2 + 2xy + 2y^2 = 1 \).
- Expressions of the form \( Ax^2 + Bxy + Cy^2 \) are called **quadratic forms**, and they don’t always represent ellipses (the e-values of the associated symmetric matrix might not be both positive). But it may still be useful to find coordinates \((u,v)\) under which the expression transforms into an expression not containing a \(uv\) term. The procedure is the same as the above. Do that for \( x^2 - 12xy - 4y^2 = 1 \).

**Justification.** The connection with matrices is in the fact that we can always write an expression of the form

\[ ax^2 + 2bxy + cy^2 \]

using vectors and a symmetric matrix:

\[
\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\
    b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}
\]

(verify this; the triple matrix product yields a \( 1 \times 1 \) matrix whose only entry is precisely \( ax^2 + 2bxy + cy^2 \)). Let us call

\[
x = \begin{bmatrix} x \\ y \end{bmatrix}, \quad A = \begin{bmatrix} a & b \\
    b & c \end{bmatrix}
\]

We just learned that, because \( A \) is symmetric, there exist a diagonal matrix \( D \) and an orthogonal matrix \( P \) such that \( A = PDP^T \). Inserting this into the expression \( x^TAx \) above, we obtain

\[
x^T(PDP^T)x = (P^Tx)^TD(P^Tx)
\]

If we denote the entries of \( P^Tx \) and \( D \) as

\[
P^Tx = \begin{bmatrix} u \\ v \end{bmatrix}, \quad D = \begin{bmatrix} \lambda_1 & 0 \\
    0 & \lambda_2 \end{bmatrix}
\]

we can see that this expression will become

\[ \lambda_1 u^2 + \lambda_2 v^2 \]

(this is similar to the above calculation). And since

\[
\begin{bmatrix} u \\ v \end{bmatrix} = P^T \begin{bmatrix} x \\ y \end{bmatrix} = P^{-1} \begin{bmatrix} x \\ y \end{bmatrix}
\]

the change-of-coordinates matrix from the \((x,y)\) system to the \((u,v)\) system is \( P \), which is a \( 2 \times 2 \) orthogonal matrix. If we managed to select \( P \) having determinant \( 1 \), it must be a rotation matrix:

\[
P = \begin{bmatrix} \cos \theta & -\sin \theta \\
    \sin \theta & \cos \theta \end{bmatrix}
\]
Suggested problems from the book

- **Systematic**: 1-10 (but ignore item (e)), 13-20
- **True or false**: 21-25, 28-31, 33-35, 39
- **Conceptual**: 41