Advanced Engineering Mathematics Lecture Notes

West Virginia Wesleyan College

Advanced Engineering Mathematics Lecture Notes

West Virginia Wesleyan College

Jesse Oldroyd
West Virginia Wesleyan College

March 18, 2022

Edition: Annual Edition 2022

Website: j-oldroyd.github.io/wvwc-advanced-engineering-math/output/html/wvwc-advanced-engi

Acknowledgements

Preface

This document was created to serve as a set of lecture notes for Advanced Engineering Mathematics (MATH 301) at West Virginia Wesleyan College. The goal of this course is to become familiar with basic concepts in linear algebra and multivariable calculus. As such, the course is naturally divided into two more or less independent parts: Part I and Part II.

This document also endeavors to include many examples that use code via Sage Cells (`cloud.sagemath.com`), since many computations in linear algebra and multivariable calculus are tedious at best to do by hand. The programming languages used are Octave (`www.gnu.org/software/octave/index`) and Sage (`sagemath.org`), which are both free and open source systems. Familiarity with these programming languages is neither assumed nor required for this course.

Contents

Part I

Linear Algebra

Chapter 1

Introduction to Matrix Algebra

1.1 Matrices, Vectors and Linear Combinations

The primary objects of study in the field of linear algebra and its applications are linear transformations between vector spaces. These linear transformations are often represented using *matrices*.

Definition 1.1 Matrix. A **matrix** is a rectangular array of numbers. If this array has m rows and n columns, we say the matrix is an $m \times n$ matrix. $\lozenge$

The following are examples of matrices:

$$\begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & -2 & 0 \\ 3 & -21 & 2 \end{bmatrix}$$

The first is 2×2 and the second is 2×3.

We say that a matrix is a **square matrix** if it has the same number of rows as columns.

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

The **diagonal entries** are $a_{ii}, 1 \leq i \leq n$ and these form the **main diagonal** of the matrix.

As important as matrices are in applications of mathematics, many computing solutions exist for handling computations involving them. One open source solution (included in Sage/CoCalc!) is Octave[1], which is a free alternative to MATLAB. In the code cell below Octave is used to define the square matrix above and get its diagonal entries. Note that brackets must be used to contain the entries of the matrix, entries in the same row must be separated by commas (or spaces) and rows are separated by semicolons.

```
A = [1, 2; -3, 4]
diag(A)
```

[1] https://www.gnu.org/software/octave/index

1.2 Matrix Multiplication

To be completed.

1.3 Systems of Linear Equations

To be completed.

1.4 Linear Independence

To be completed.

Bases in $\mathbb{R}^n$

Definition 1.2 Basis of $\mathbb{R}^n$. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ denote a subset of vectors in $\mathbb{R}^n$. We say that $\mathcal{B}$ is a **basis** for $\mathbb{R}^n$ if it satisfies the following properties:

1. $\mathcal{B}$ is linearly independent.
2. $\mathcal{B}$ is a spanning set, i.e., $\operatorname{span} \mathcal{B} = \mathbb{R}^n$.

$\Diamond$

Definition 1.3 Pivot Columns. Let A be a matrix. The **pivot columns** of A are those columns which contain leading entries in any echelon form of A. $\Diamond$

Theorem 1.4 Rank and Pivot Columns. *Let A be a matrix. Then* $\operatorname{rank} A$ *is exactly equal to the number of pivot columns of A.*

Definition 1.5 Column Space. The **column space** of a matrix A is the span of the columns of A. Equivalently, the column space is the set of all vectors of the form $A\mathbf{x}$. The column space of A is denoted by $\operatorname{col} A$. $\Diamond$

1.5 Existence of Solutions

Recall that any linear system may be expressed as a matrix equation of the form $A\mathbf{x} = \mathbf{b}$. Based on our previous work, we can make the following observations.

Theorem 1.6 Consistency, Rank and the Column Space. *Let A be an $m \times n$ matrix, $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$. The linear system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b} \in \operatorname{col} A$. Equivalently, the system is consistent if and only if $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & \mathbf{b} \end{bmatrix}$. Furthermore, the system has precisely one solution if $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & \mathbf{b} \end{bmatrix} = n$ and has infinitely many solutions if $\operatorname{rank} A = \operatorname{rank} \begin{bmatrix} A & \mathbf{b} \end{bmatrix} < n$.*

As before, we use Gaussian elimination (i.e., row reduction) to solve systems.

Example 1.7 Find any solutions of

$$\begin{aligned} 5x - 7y + 3z &= 17 \\ -15x + 20y - 9z &= -50 \end{aligned}$$

Solution. Reducing to an echelon form is enough to determine if the system is consistent and find the number of solutions it has. Using Octave to find the reduced echelon form (see the code cell immediately after this example), we get

$$\begin{bmatrix} 1 & 0 & \frac{3}{5} & 2 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

From the reduced echelon form above, we see that the system must be consistent since the rank of the coefficient matrix is equal to the rank of the augmented matrix. Equivalently, the last column is not a pivot column. Furthermore, there are infinitely many solutions since the rank of the coefficient matrix is less than the total number of columns.

The solution set itself can be written in vector notation as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -\frac{3}{5} \\ 0 \\ 1 \end{bmatrix}.$$

This, again, is verified below. □

```
A = [5, -7, 3; -15, 20, -9];
b = [17; -50];
rref([A, b]) % reduced echelon form of augmented matrix

% verify solution
z = 3.1; % arbitrary value for z
x_soln = [2; -1; 0] + z*[-3/5; 0; 1]; % solution vector
A*x_soln % equals b
```

In the last example the variable z led to infinitely many solutions and we were able to write out solution depending on the value of this variable. We call z a *free variable* and x and y *basic variables*.

Definition 1.8 Basic and Free Variables. Given a consistent linear system $A\mathbf{x} = \mathbf{b}$, the variables corresponding to pivot columns of A are **basic variables** and the variables corresponding to non-pivot columns of A are **free variables**. ◊

Any solution of the linear system $A\mathbf{x} = \mathbf{b}$ can always be written to depend solely on any free variables as we did in Example 1.7. In fact we can go a bit further, still using our answer in Example 1.7 as a guide. Using free variables, any solution to $A\mathbf{x} = \mathbf{b}$ can be written as a sum of two components:

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_{\text{free}}.$$

This notation will change shortly, but the main idea is that one component of the solution will not depend on the free variable and will represent a single solution of the system $A\mathbf{x} = \mathbf{b}$. In Example 1.7 this would be

$$\mathbf{x}_p = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix},$$

and it's easy to verify that

$$A\mathbf{x}_p = \begin{bmatrix} 5 & -7 & 3 \\ -15 & 20 & -9 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 17 \\ -50 \end{bmatrix}.$$

The *other* component of the solution, $\mathbf{x}_{\text{free}}$, will depend on the free variable. In Example 1.7, this component was

$$\mathbf{x}_{\text{free}} = z \begin{bmatrix} -\frac{3}{5} \\ 0 \\ 1 \end{bmatrix}.$$

As it turns out, this component is *not* a solution of the original system $A\mathbf{x} = \mathbf{b}$. Instead, $A\mathbf{x}_{\text{free}} = \mathbf{0}$. This behavior is shared by all consistent systems with free variables, and leads us to introduce the following terminology.

Definition 1.9 Associated Homogeneous System. Given a linear system $A\mathbf{x} = \mathbf{b}$, we define the **associated homogeneous system** to be the system $A\mathbf{x} = \mathbf{0}$. ♢

The observations made after Example 1.7 can be summarized in the following theorem.

Theorem 1.10 Particular and Homogeneous Solutions. *Suppose that $A\mathbf{x} = \mathbf{b}$ is a consistent linear system. Then the general solution $\mathbf{x}$ can be written in the form $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$ where $\mathbf{x}_p$ is a single solution of the original system $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x}_h$ is the general solution of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$. We call $\mathbf{x}_p$ a **particular solution**.*

Proof. Here, we only prove that $\mathbf{x}_h$ satisfies the associated homogeneous system. Since $\mathbf{x}_p$ is a solution of the original system along with $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, it follows that

$$A\mathbf{x}_h = A(\mathbf{x} - x_p) = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

Therefore $\mathbf{x}_h$ is a solution of the associated homogeneous system. ■

Since solutions of homogeneous systems play such an important role in the solution of non-homogeneous systems, we give their solution sets a special name.

Definition 1.11 Null Space. Let A be a matrix. The **null space** of A is the set of all solutions of $A\mathbf{x} = \mathbf{0}$. This is denoted by $\operatorname{null} A$. ♢

As with column spaces and row spaces, null spaces are always subspaces.

Theorem 1.12 The Null Space is a Subspace. *Let A be an $m \times n$ matrix. Then $\operatorname{null} A$ is a subspace of $\mathbb{R}^n$.*

Proof. To show that $\operatorname{null} A$ is a subspace we need to show that it's closed under linear combinations. So let $\mathbf{u}, \mathbf{v} \in \operatorname{null} A$ be arbitrary vectors in the null space and let $\alpha, \beta \in \mathbb{R}$ be arbitrary scalars. Our goal is to show that $\alpha\mathbf{u} + \beta\mathbf{v} \in \operatorname{null} A$. Thankfully, we can do this very quickly:

$$\begin{aligned} A(\alpha\mathbf{u} + \beta\mathbf{v}) &= \alpha A\mathbf{u} + \beta A\mathbf{v} \\ &= \mathbf{0} + \mathbf{0} \\ &= \mathbf{0}, \end{aligned}$$

which shows that the linear combination $\alpha\mathbf{u} + \beta\mathbf{v}$ lies in $\operatorname{null} A$. ■

The concept of the null space is related to that of the column space in Definition 1.5, but they are distinct. To be precise, if A is an $m \times n$ matrix then

$$\begin{aligned} \operatorname{col} A &= \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m \\ \operatorname{null} A &= \{\mathbf{x} : A\mathbf{x} = \mathbf{0}\} \subseteq \mathbb{R}^n \end{aligned}$$

Example 1.13 Finding a Null Space. Let

$$A = \begin{bmatrix} 0 & 5 & 5 & -10 & 0 \\ 2 & -3 & -3 & 6 & 2 \\ 4 & 1 & 1 & -2 & 4 \end{bmatrix}.$$

Find $\operatorname{null} A$.

Solution. We need to find the solution set of $A\mathbf{x} = \mathbf{0}$ which we've done before. The Octave code cell below can be used to solve this system, giving the reduced echelon form for the augmented matrix $\begin{bmatrix} A & \mathbf{0} \end{bmatrix}$ to be

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Therefore any solution $\mathbf{x} \in \mathbb{R}^5$ of $A\mathbf{x} = \mathbf{0}$ must look like

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -x_5 \\ -x_3 + 2x_4 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Note that the above shows that

$$\operatorname{null} A = \operatorname{span} \left\{ \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Since these vectors are also linearly independent, it follows that the set

$$\left\{ \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is in fact a basis for $\operatorname{null} A$ and $\dim \operatorname{null} A = 3$. □

```
% code cell for finding null space in previous example
```

At this point we can make a simple but useful observation. In Example 1.13, the dimension of the null space was directly tied to the number of free variables in the system $A\mathbf{x} = \mathbf{0}$. The number of basic variables is likewise equal to the number of pivot columns of A. Noting that the number of basic variables plus the number of free variables must be the total number of columns of A, together with Theorem 1.4, we get the *Rank-Nullity Theorem*.

Theorem 1.14 Rank-Nullity Theorem. *Let A be an $m \times n$ matrix. Then*

$$\operatorname{rank} A + \dim \operatorname{null} A = n.$$

1.6 Determinants

If A is an $n \times n$ square matrix and if $\mathbf{x} \in \mathbb{R}^n$, then $\mathbf{x}$ and $A\mathbf{x}$ have the same size. In other words, both $\mathbf{x}$ and $A\mathbf{x}$ live in the same vector space $\mathbb{R}^n$. This makes some geometry involving A slightly easier. In particular, $\operatorname{col} A$ must be a subspace of $\mathbb{R}^n$.

If $\operatorname{col} A$ is a subspace of $\mathbb{R}^n$ then we can say that the linear system $A\mathbf{x} = \mathbf{b}$ is always consistent if and only if $\operatorname{col} A = \mathbb{R}^n$. If this were not the case, then there would exist some vector $\mathbf{b} \in \mathbb{R}^n$ such that $\mathbf{b} \notin \operatorname{col} A$ and so $A\mathbf{x} = \mathbf{b}$ would have to be inconsistent. So square matrices A for which $\operatorname{col} A = \mathbb{R}^n$ are particularly well-behaved and useful. Therefore we'd like to develop conditions to check for when a square matrix A satisfied this property.

One possible condition for this is the following: $\operatorname{col} A = \mathbb{R}^n$ if and only if $\dim \operatorname{col} A = n$, which happens if and only if $\operatorname{rank} A = n$. Therefore $A\mathbf{x} = \mathbf{b}$ is always consistent if and only if $\operatorname{rank} A = n$. However, another useful condition which uses geometry is the following: $\operatorname{col} A = \mathbb{R}^n$ if and only if the columns of A span an n-dimensional figure in $\mathbb{R}^n$. To get an idea of why this should be true, consider a 2×2 matrix A whose columns determine a parallelogram (as opposed to a line) in $\mathbb{R}^2$ such as

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}.$$

Since $\operatorname{col} A$ is the set of all linear combinations of columns of A, and geometrically this is just the set of all points that we can reach in $\mathbb{R}^2$ by stretching and expanding the parallelogram determined by the columns, it follows that $\operatorname{col} A = \mathbb{R}^2$.

The *determinant* makes this observation precise. Given an $n \times n$ square matrix A, the determinant of A represents the (signed) volume of the parallelepiped determined by the columns of A. If this volume is nonzero then this means that the parallelepiped must be an n-dimensional figure in $\mathbb{R}^n$ and so the column space of A would be all of $\mathbb{R}^n$. In the 2×2 case it's not too difficult to compute the determinant. If

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

then $\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$. Note that $ad - bc$ does give the area of the parallelogram determined by the columns of A. In three dimensions and higher the formula becomes more complicated and must be defined recursively.

Definition 1.15 Determinant of a Matrix. Let $A = [\, a_{ij} \,]$ be an $n \times n$ matrix. Let A_{ij} denote the *sub-matrix* of A obtained by removing the i^{th} row and j^{th} column of A (the same row and column containing the entry a_{ij}). Then the **determinant** of A is defined recursively by the formula

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}).$$

We'll see later that we can expand along any row or column, but for now we'll stick to the first row.

◊

Example 1.16 Computing a Determinant. Let

$$A = \begin{bmatrix} 2 & -6 & 4 \\ 3 & 5 & -2 \\ 1 & 6 & 3 \end{bmatrix}.$$

Find $\det(A)$.

Solution. The formula in Definition 1.15 states that

$$\det(A) = 2\begin{vmatrix} 5 & -2 \\ 6 & 3 \end{vmatrix} - (-6)\begin{vmatrix} 3 & -2 \\ 1 & 3 \end{vmatrix} + 4\begin{vmatrix} 3 & 5 \\ 1 & 6 \end{vmatrix}$$

which simplifies to 172. This can be confirmed in the Octave cell below. □

```
A = [2, -6, 4; 3, 5, -2; 1, 6, 3]
det(A)
```

When computing determinants by hand, it's often useful to expand along the row or column containing the most zeros instead of just the first row. As long as we're careful about signs, the next result says this is permissible.

Theorem 1.17 Cofactor Expansion. *Let $A = [a_{ij}]$ be an $n \times n$ matrix and define A_{ij} as in Definition 1.15. Then*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

Example 1.18 Computing a Determinant with a Cofactor Expansion. Let

$$A = \begin{bmatrix} 1 & -2 & 5 & 2 \\ 0 & 0 & 3 & 0 \\ 2 & -6 & -7 & 5 \\ 5 & 0 & 4 & 4 \end{bmatrix}.$$

Find $\det(A)$.

Solution. We can save some work by expanding along the second row to take advantage of the zeros that appear. Doing so, we get

$$\det(A) = -3 \begin{vmatrix} 1 & -2 & 2 \\ 2 & -6 & 5 \\ 5 & 0 & 4 \end{vmatrix} = -3 \left(5 \begin{vmatrix} -2 & 2 \\ -6 & 5 \end{vmatrix} + 4 \begin{vmatrix} 1 & -2 \\ 2 & -6 \end{vmatrix} \right)$$

□

Computing determinants becomes very simple when working with *triangular matrices*.

Definition 1.19 Triangular Matrices. A matrix A is **lower** (respectively, **upper**) **triangular** is all of the entries above (respectively, below) the main diagonal are 0. A matrix is **triangular** if it is lower triangular or upper triangular. ♢

Example 1.20 Computing the Determinant of a Triangular Matrix. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & -2 \end{bmatrix}.$$

Find $\det(A)$.

Solution. Using appropriate cofactor expansions, we see that

$$\det(A) = 1 \cdot 4 \cdot (-2).$$

□

Theorem 1.21 Determinants of Triangular Matrices. *Let A be an $n \times n$ triangular matrix. Then $\det(A)$ is just the product of its diagonal entries:*

$$\det(A) = \prod_{i=1}^{N} a_{ii}.$$

Theorem 1.21 leads to another approach for finding determinants via row reduction. If we have a square matrix A and we can reduce it to echelon form, then it becomes very easy to find the determinant of the echelon form. If we can then relate this determinant back to $\det(A)$, then we would be able to find $\det(A)$ using the echelon form instead. It turns out this can be done as follows.

Theorem 1.22 Row Operations and the Determinant. *Let A and B denote square matrices of the same size and suppose that B is obtained from A by performing a single row operation.*

1. *If the row operation was row replacement, then* $\det(A) = \det(B)$.
2. *If the row operation was row scaling by a factor of k, then* $\det(B) = k\det(A)$.
3. *If the row operation was row interchange, then* $\det(B) = -\det(A)$.

Two other useful results about determinants are given below.

Theorem 1.23 Multiplicative Property. *Let A and B denote square matrices of the same size. Then* $\det(AB) = \det(A)\det(B)$.

Theorem 1.24 Determinants and the Transpose. *Let A be a square matrix. Then* $\det(A) = \det\left(A^T\right)$.

1.7 Matrix Inverses

Consider the equation $75x = 2$. We can solve this quite easily for x by dividing both sides by 75, or equivalently, multiplying both sides of the equation by $frac175 = 75^{-1}$. 75^{-1} is the *multiplicative inverse* of the number 75, and so when multiplied to it we are left with only the number 1. We want to do the same with the matrix equation $A\mathbf{x} = \mathbf{b}$; that is, we want to find an *inverse matrix* A^{-1} that, when multiplied to A, leaves only the identity matrix.

Definition 1.25 Invertible Matrices. An $n \times n$ matrix A is said to be **invertible** (or **nonsingular**) if there exists a matrix A^{-1} such that $A^{-1}A = AA^{-1} = I_n$. We call A^{-1} the **inverse** of A. If a matrix is *not* invertible, then we say that it is **singular**. $\Diamond$

Note that if A is a square matrix and C is another square matrix such that either $AC = I$ or $CA = I$, then $C = A^{-1}$.

Example 1.26 Confirming a Matrix Inverse. Let

$$A = \begin{bmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{bmatrix}.$$

1. Show that the matrix
$$C = \begin{bmatrix} 8 & 3 & 1 \\ 10 & 4 & 1 \\ \frac{7}{2} & \frac{3}{2} & \frac{1}{2} \end{bmatrix}$$
is the inverse of A.

2. Let
$$\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$
Solve $A\mathbf{x} = \mathbf{b}$.

Solution.

1. All we need to do is to show that $AC = I$. This can be done quickly using Octave as in the code cell below.

2. The solution is $\mathbf{x} = A^{-1}\mathbf{b}$. Given that we now know A^{-1}, we can solve this quickly.

□

```
A = [1, 0, -2; -3, 1, 4; 2, -3, 4];
C = [8,3,1; 10,4,1; 7/2, 3/2, 1/2];

A*C % identity matrix

b = [1;1;1]
x = C*b % x = inv(A)*b is the solution
```

Example 1.27 Inverse of an Orthogonal Matrix. Let U be an orthogonal matrix. What is U^{-1}?

Solution. $U^{-1} = U^T$. So it is *very* easy to find the inverse of an orthogonal matrix. □

An important property about determinants is that they say precisely when a matrix is invertible. If A is an $n \times n$ matrix, then A has an inverse if and only if $\det A \neq 0$.

Theorem 1.28 Invertibility and Solutions of Systems. *Let A be an invertible $n \times n$ matrix. Then for each $\mathbf{b} \in \mathbb{R}^n$, the matrix equation $A\mathbf{x} = \mathbf{b}$ has* exactly *one solution: $\mathbf{x} = A^{-1}\mathbf{b}$.*

Proof. To prove this statement we must show two things:

1. $A^{-1}\mathbf{b}$ is a solution.

2. $A^{-1}\mathbf{b}$ is the only solution.

We start with the first item. To check that $A^{-1}\mathbf{b}$ is a solution of $A\mathbf{x} = \mathbf{b}$, we just plug it in for $\mathbf{x}$ and simplify:

$$A(A^{-1}\mathbf{b}) = I\mathbf{b} = \mathbf{b}.$$

Hence this is a solution.

To show that this is the only solution, suppose that $\mathbf{u}$ is some other solution of $A\mathbf{x} = \mathbf{b}$. We must show that $\mathbf{u} = A^{-1}\mathbf{b}$. Since $\mathbf{u}$ is assumed to be a solution, we have

$$\mathbf{u} = \mathbf{b} \Rightarrow \mathbf{u} = A^{-1}\mathbf{b}.$$

Hence $A^{-1}\mathbf{b}$ is the only solution. ■

Computing the Inverse of a Matrix

For 2×2 matrices, we have a simple formula for the inverse.

Theorem 1.29 Inverse of 2×2 Matrices. *Let*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

If $ad - bc \neq 0$, then

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

The quantity $ad - bc$ in Theorem 1.29 can also be recognized as the determinant of the matrix A. See Definition 1.15.

Example 1.30 Show that the system

$$\begin{aligned} 9x_1 + 3x_2 &= -9 \\ -6x_1 - 3x_2 &= 4 \end{aligned}$$

is consistent and then find the solution.

Solution. We can rewrite this as the matrix equation $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 9 & 3 \\ -6 & -3 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} -9 \\ 4 \end{bmatrix}.$$

We can show that the system is consistent by computing $\det A$: since $\det A = -9 \neq 0$, A^{-1} exists. And since A^{-1} exists, the system must be solvable.

To solve it, we use the above formula to compute A^{-1}:

$$A^{-1} = -\frac{1}{9}\begin{bmatrix} -3 & -3 \\ 6 & 9 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & -1 \end{bmatrix}.$$

So the (unique) solution is

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & -1 \end{bmatrix}\begin{bmatrix} -9 \\ 4 \end{bmatrix}.$$

□

Theorem 1.31 *Let A and B be invertible $n \times n$ matrices.*

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

The Invertible Matrix Algorithm

We can now find the inverse of a 2×2 matrix; we need to determine how to find the inverse of a larger matrix. To do this, we will use *elementary matrices*.

Definition 1.32 Elementary Matrices. An **elementary matrix** is a matrix obtained by performing a single elementary row operation on the identity matrix. ♢

Example 1.33 The matrices

$$E_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_2 = \begin{bmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } E_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

are elementary matrices. The first corresponds to scaling the first row by 2; the second corresponds to adding five times the second row to the first row; and the third corresponds to switching rows one and three. □

The important fact about elementary matrices is that multiplying them to *any* matrix has the same effect as performing the corresponding elementary row operation on the matrix.

Example 1.34 Let

$$A = \begin{bmatrix} 1 & 2 & 9 \\ 0 & 3 & 3 \\ 4 & 4 & 1 \end{bmatrix}.$$

Use elementary matrices to perform the following row operations:

1. Add two times row three to row two.
2. Scale row three by -3.
3. Swap row two with row one and then add five times row three to row one.

Solution. For each case, we only need to determine the elementary matrix corresponding to each row operation. The elementary matrix for the first operation is

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}.$$

To perform this operation on A, we just multiply E_1 and A:

$$\begin{aligned} A &\overset{2R_3+R_2}{\sim} E_1A \\ &= \begin{bmatrix} 1 & 2 & 9 \\ 8 & 11 & 5 \\ 4 & 4 & 1 \end{bmatrix} \end{aligned}$$

which matches with the matrix we would have obtained just using a row operation.

The elementary matrix we need for the next operation is

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{bmatrix}$$

and so

$$A \overset{-3R_3}{\sim} E_2A.$$

Finally, we have two elementary row operations here, so we can't just use a single elementary matrix. We'll need to use two; one for each row operation:

$$E_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E_4 = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

So

$$A \underset{R_1 \leftrightarrow R_2}{\overset{5R_3+R_1}{\sim}} E_4E_3A.$$

□

So row operations on a matrix can be viewed as multiplications by elementary matrices. And since row operations are invertible, elementary matrices are invertible as well. To find the inverse of an elementary matrix E, just write down the elementary matrix corresponding to the row operation that transforms E back into I.

Example 1.35 Inverse of an Elementary Matrix. Let E_1, E_2 and E_4 be as above. Find the inverse of each matrix.

Solution. We have

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}, \qquad E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{3} \end{bmatrix} \quad \text{and} \quad E_4^{-1} = \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

□

Theorem 1.36 Invertible Matrix Algorithm. *Let A be an $n \times n$ matrix. If $A \sim I$, then A is invertible.*

Proof. Suppose that A is row equivalent to the identity matrix I. Then we can find elementary matrices $E_1, E_2, \ldots, E_p$ such that

$$I = E_p E_{p-1} \cdots E_2 E_1 A.$$

Since elementary matrices are invertible, their product must be as well. So we can write

$$(E_p \cdots E_1)^{-1} = A.$$

Since A is the inverse of an invertible matrix, it must itself be invertible and furthermore

$$A^{-1} = E_p \cdots E_1.$$

■

The above theorem tells us that the sequence of row operations that reduces A to I also turns I into A^{-1}. This gives us an algorithm for finding the inverse of a matrix. We show this with an example.

Example 1.37 Let

$$A = \begin{bmatrix} -1 & -7 & -3 \\ 2 & 15 & 6 \\ 1 & 3 & 2 \end{bmatrix}.$$

Compute A^{-1}.

Solution. We set up the augmented matrix $\begin{bmatrix} A & I \end{bmatrix}$. The algorithm works by finding the reduced echelon form; the resulting augmented matrix is then $\begin{bmatrix} I & A^{-1} \end{bmatrix}$.

$$\begin{bmatrix} -1 & -7 & -3 & 1 & 0 & 0 \\ 2 & 15 & 6 & 0 & 1 & 0 \\ 1 & 3 & 2 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 7 & 3 & -1 & 0 & 0 \\ 2 & 15 & 6 & 0 & 1 & 0 \\ 1 & 3 & 2 & 0 & 0 & 1 \end{bmatrix}$$

$$\underset{-2R_1+R_2}{\overset{-R_1+R_3}{\sim}} \begin{bmatrix} 1 & 7 & 3 & -1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & -4 & -1 & 1 & 0 & 1 \end{bmatrix}$$

$$\overset{4R_2+R_3}{\sim} \begin{bmatrix} 1 & 7 & 3 & -1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & -1 & 9 & 4 & 1 \end{bmatrix}$$

$$\overset{3R_3+R_1}{\sim} \begin{bmatrix} 1 & 7 & 0 & 26 & 12 & 3 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & -1 & 9 & 4 & 1 \end{bmatrix}$$

$$\underset{-7R_2+R_1}{\overset{-R_3}{\sim}} \begin{bmatrix} 1 & 0 & 0 & 12 & 5 & 3 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & -9 & -4 & -1 \end{bmatrix}$$

So

$$A^{-1} = \begin{bmatrix} 12 & 5 & 3 \\ 2 & 1 & 0 \\ -9 & -4 & -1 \end{bmatrix}.$$

□

Example 1.38 A square matrix A of size 10×10. Suppose that A has rank 9. Is A invertible?

Solution. No! This is because A does not have a pivot in each row (since rank $A = 9$, A only has 9 pivots). Therefore we can't row reduce A to get I. Since A is not row equivalent to the identity matrix, A cannot be invertible. □

1.8 *LU* Decomposition

The approach we used in Theorem 1.36 also gives us a useful way to *factor* matrices. As with polynomials, the goal in factoring a matrix A is to write it as the product of simpler matrices. The factorization we'll look at in this section is the LU decomposition, which we demonstrate first by an example.

Example 1.39 ***LU* Factorization.** Let

$$A = \begin{bmatrix} 3 & 2 & -1 \\ -2 & 4 & 4 \\ 3 & 1 & 2 \end{bmatrix}.$$

Now consider the problem of reducing A to an echelon form. First, we can take $\frac{2}{3}$ times row 1 and add it to row 2 and -1 times row 1 and add it to row 3 to obtain

$$\begin{bmatrix} 3 & 2 & -1 \\ 0 & \frac{16}{3} & \frac{10}{3} \\ 0 & -1 & 3 \end{bmatrix}.$$

A final operation, adding $\frac{3}{16}$ times row 2 to row 3, gives us the echelon form

$$U = \begin{bmatrix} 3 & 2 & -1 \\ 0 & \frac{16}{3} & \frac{10}{3} \\ 0 & 0 & \frac{29}{8} \end{bmatrix}$$

Now, we also saw in the last section that row operations correspond to multiplication by elementary matrices. In this case, $U = E_{23}E_{13}E_{12}A$ where

$$E_{12} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{2}{3} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_{13} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \text{ and } E_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{3}{16} & 1 \end{bmatrix}.$$

If we now solve the equation $U = E_{23}E_{13}E_{12}A$ for A, we get the useful product

$$A = E_{12}^{-1}E_{13}^{-1}E_{23}^{-1}U.$$

It turns out to be very easy to find $E_{12}^{-1}E_{13}^{-1}E_{23}^{-1}$ as these are elementary matrices, and we get

$$E_{12}^{-1}E_{13}^{-1}E_{23}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{2}{3} & 1 & 0 \\ 1 & -\frac{3}{16} & 1 \end{bmatrix}$$

This matrix is lower triangular, and we denote it by L. Thus we have the factorization $A = LU$. □

We can verify the factorization in Example 1.39 using Octave as below:

```
A = [3,2,-1;-2,4,4;3,1,2];
L = [1,0,0;-2/3,1,0;1,-3/16,1];
U = [3,2,-1;0,16/3,10/3;0,0,29/8];
L*U
```

The Octave command `lu()` can also be used to find matrices L and U satisfying $A = LU$:

```
% run the previous cell first!
[L1, U1] = lu(A);
L1*U1
```

The matrix factorization $A = LU$ demonstrated above can always be computes (reordering the rows of A if necessary). In this factorization, L is lower triangular with 1s along the diagonal and U is an echelon form of A (and therefore upper triangular if A is square). This form has several advantages. First, if A is square then $\det(A) = \det(L)\det(U) = \det(U)$, and the determinant of U is very easy to find. Second, this form lets us solve systems more quickly.

Example 1.40 Solving a System with LU factorization. Let

$$A = \begin{bmatrix} -1 & -7 & -3 \\ 2 & 15 & 6 \\ 1 & 3 & 2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 1 \\ -5 \\ 11 \end{bmatrix}.$$

Find an LU decomposition of A and use it to solve $A\mathbf{x} = \mathbf{b}$.

Solution. We can find L and U by reducing A to an echelon form. One possibility is

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & -4 & 1 \end{bmatrix} \text{ and } U = \begin{bmatrix} -1 & -7 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Now we can solve $A\mathbf{x} = \mathbf{b}$ in two steps:

1. Solve $L\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$.
2. Solve $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$.

Solving the first equation by reducing $[\, L \; \mathbf{b} \,]$ produces

$$\mathbf{y} = \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix}.$$

Then, solving the second equation by reducing $[\, U \; \mathbf{y} \,]$ produces

$$\mathbf{x} = \begin{bmatrix} 20 \\ -3 \\ 0 \end{bmatrix}.$$

□

The process in Example 1.40 produces the same exact answer that row reduction would if applied to $[\, A \; \mathbf{b} \,]$, or if A^{-1} were found and multiplied to $\mathbf{b}$. However, computing and using the LU factorization can be more stable numerically than either Gaussian elimination or computing the inverse.

Chapter 2

Eigenvalues and Eigenvectors

2.1 Finding Eigenvalues and Eigenvectors

Many applications call for computing matrix-vector products like $A\mathbf{x}$, and in such cases it often happens that A is a square matrix. If many such products need to be computed, it'd be nice to know if there was a way to simplify these calculations. One possible way to approach this is by using *eigenvectors and eigenvalues*.

Definition 2.1 Eigenvalues and Eigenvectors. An **eigenvector** of an $n \times n$ matrix A is a *nonzero vector* $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a *nonzero* solution of the equation $A\mathbf{x} = \lambda\mathbf{x}$. ◊

Note that the zero vector is not allowed to be an eigenvector, but zero is allowed to be an eigenvalue.

Example 2.2 Verifying Eigenvectors. Let

$$A = \begin{bmatrix} 5 & 2 \\ 3 & 6 \end{bmatrix}, \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_3 = \mathbf{0}.$$

Which of the vectors, if any, is an eigenvector of A? What is one eigenvalue of A?

Solution. First, note that $\mathbf{v}_3 = \mathbf{0}$ is not an eigenvector since it is the zero vector. So we'll check if the other two vectors are eigenvectors:

$$A\mathbf{v}_1 = \begin{bmatrix} 3 \\ -3 \end{bmatrix} = 3\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and

$$A\mathbf{v}_2 = \begin{bmatrix} 11 \\ 12 \end{bmatrix}.$$

So $\mathbf{v}_1$ is an eigenvector of A (with eigenvalue 3), but $\mathbf{v}_2$ is not an eigenvector since there is no scalar we can multiply $\mathbf{v}_2$ by to get $A\mathbf{v}_2$. □

It's a little harder to verify if a given number is an eigenvalue of a matrix A.

Example 2.3 Verifying Eigenvalues. Is -2 an eigenvalue of the matrix

$$Q = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}?$$

If it is, find a corresponding eigenvector.

Solution. -2 is an eigenvalue of Q if and only if the equation $Q\mathbf{x} = -2\mathbf{x}$ has a *nonzero* solution. Rearranging this equation, we can say that -2 is an eigenvalue of Q if and only if $(Q + 2I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. So we'll row reduce the augmented matrix $\begin{bmatrix} Q + 2I & \mathbf{0} \end{bmatrix}$ to see if the system has free variables:

$$\begin{aligned} \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \end{bmatrix} &\sim \begin{bmatrix} -1 & 2 & -1 & 0 \\ 2 & -1 & -1 & 0 \\ -1 & -1 & 2 & 0 \end{bmatrix} \\ &\underset{2R_1+R_2}{\overset{-R_1+R_3}{\sim}} \begin{bmatrix} -1 & 2 & -1 & 0 \\ 0 & 3 & -3 & 0 \\ 0 & -3 & 3 & 0 \end{bmatrix} \\ &\sim \begin{bmatrix} -1 & 2 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

so there are free variables and it follows that -2 is indeed an eigenvalue of Q.

To find an eigenvector, we just need to find a nontrivial solution of $(Q + 2I)\mathbf{x} = \mathbf{0}$---equivalently, a nonzero vector in $\text{null}\,(Q + 2I)$---so we'll continue row reducing:

$$\begin{aligned} \begin{bmatrix} -1 & 2 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} &\overset{-2R_2+R_1}{\sim} \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ &\sim \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

So

$$\text{null}\,(Q + 2I) = \text{span} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

and a single eigenvector of Q is given by

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

□

Computing Eigenvalues and Eigenvectors

There are two things we can note from Example 2.3. First, any nonzero vector in $\text{null}\,(Q + 2I)$ is an eigenvector of Q with eigenvalue -2. This leads to the following definition.

Definition 2.4 Eigenspaces. Let A be an $n \times n$ matrix and suppose that λ is an eigenvalue of A. The **eigenspace** of A corresponding to λ is the subspace of $\mathbb{R}^n$ containing all of the eigenvectors corresponding to λ in addition to the zero vector. In other words, the eigenspace of A corresponding to λ is the set $\text{null}\,(A - \lambda I)$. This set is often denoted E_λ. ◊

The second item we note is that -2 was an eigenvalue precisely because the equation $(Q+2I)\mathbf{x} = \mathbf{0}$ had nontrivial solutions. In other words, $Q+2I$ *was not invertible*. In general, the polynomial $\det(A - \lambda I)$ and equation $\det(A - \lambda I) = 0$

are important enough that they deserve their own names.

Definition 2.5 Characteristic Polynomial and Characteristic Equation. Let A be an $n \times n$ matrix. The **characteristic polynomial** of A is the n^{th} degree polynomial $\det(A - \lambda I)$. The equation $\det(A - \lambda I) = 0$ is the **characteristic equation**. ♢

Theorem 2.6 Eigenvalues and the Characteristic Equation. *Let A be a square matrix. Then the eigenvalues of A are the solutions of the characteristic equation* $\det(A - \lambda I) = 0$.

Example 2.7 Finding Eigenvalues. Find the eigenvalues of the matrix

$$A = \begin{bmatrix} 4 & -2 & 3 \\ 0 & -1 & 3 \\ -1 & 2 & -2 \end{bmatrix}.$$

Solution. We need to compute the characteristic polynomial $\det(A - \lambda I)$. Now,

$$A - \lambda I = \begin{bmatrix} 4-\lambda & -2 & 3 \\ 0 & -1-\lambda & 3 \\ -1 & 2 & -2-\lambda \end{bmatrix},$$

so

$$\begin{aligned} \det(A - \lambda I) &= (4-\lambda)\begin{vmatrix} -1-\lambda & 3 \\ 2 & -2-\lambda \end{vmatrix} - \begin{vmatrix} -2 & 3 \\ -1-\lambda & 3 \end{vmatrix} \\ &= (4-\lambda)[(-1-\lambda)(-2-\lambda) - 6] - [-6 - (-3-3\lambda)] \\ &= (4-\lambda)(-4+3\lambda+\lambda^2) - (-3+3\lambda) \\ &= (4-\lambda)(\lambda+4)(\lambda-1) + 3(1-\lambda) \\ &= (\lambda-1)[(4-\lambda)(\lambda+4) - 3] \\ &= (\lambda-1)(13-\lambda^2) \end{aligned}$$

The solutions of the characteristic equation

$$(\lambda-1)(13-\lambda^2) = 0$$

are given by $\lambda = 1, \pm\sqrt{13}$. So the eigenvalues of A are $1, -\sqrt{13}$ and $\sqrt{3}$. □

Computer systems such as Sage and Octave can, naturally, find eigenvalues and eigenvectors as well. In Octave this is done with the `eig` command. If no output is specified then the command produces an array of eigenvalues, while if two outputs are specified the command produces two matrices: the first matrix is a matrix of eigenvector columns of A and the second matrix is a diagonal matrix of eigenvalues of A. See the code cell below.

```
format short
A = [4, -2, 3; 0, -1, 3; -1, 2, -2];
[U,D] = eig(A)
```

Matters are simplified greatly when finding eigenvalues of triangular matrices (see Theorem 1.21).

Example 2.8 Eigenvalues of a Triangular Matrix. Find the eigenvalues of the matrix

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 \\ -1 & -3 & -3 & 0 \\ 0 & 0 & 1 & -10 \end{bmatrix}.$$

Solution. To find the eigenvalues, we need to first find $\det(B - \lambda I)$. Since

$$B - \lambda I = \begin{bmatrix} 1-\lambda & 0 & 0 & 0 \\ -1 & \lambda & 0 & 0 \\ -1 & -3 & -3-\lambda & 0 \\ 0 & 0 & -1 & -10-\lambda \end{bmatrix}$$

is triangular (just as B is triangular), it follows that

$$\det(B - \lambda I) = (1-\lambda)\lambda(-3-\lambda)(-10-\lambda).$$

The solutions of the characteristic equation, and thus the eigenvalues of B, are given by $\lambda = 1, 0, -3, -10$. $\square$

Example 2.8 suggests the following theorem.

Theorem 2.9 Eigenvalues of a Triangular Matrix. *Let A be a square triangular matrix. Then the eigenvalues of A are just the diagonal entries of A.*

So now we have a good idea of how to find eigenvalues of a square matrix A: just solve the characteristic equation $\det(A - \lambda I) = 0$. To find the corresponding eigenvectors, we need to solve the related equation $A\mathbf{v} = \mathbf{0}$, which reduces to solving $(A - \lambda I)\mathbf{v} = \mathbf{0}$.

Example 2.10 Finding Eigenvalues and Eigenvectors. Find the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 2 & 0 & -1 \\ 0 & \frac{1}{2} & 0 \\ 1 & 0 & 4 \end{bmatrix}.$$

Solution. First, we need to find the eigenvalues. Since

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 0 & -1 \\ 0 & \frac{1}{2}-\lambda & 0 \\ 1 & 0 & 4-\lambda \end{vmatrix} = (2-\lambda)\left(\frac{1}{2}-\lambda\right)(4-\lambda) + \frac{1}{2} - \lambda$$

which simplifies to

$$\det(A - \lambda I) = \left(\frac{1}{2}-\lambda\right)[(2-\lambda)(4-\lambda)+1] = \left(\frac{1}{2}-\lambda\right)[\lambda^2 - 6\lambda + 9],$$

we see that the eigenvalues of A are given by $\lambda = \frac{1}{2}, 3$.

Now we can start trying to find eigenvectors. First, we'll row reduce $\begin{bmatrix} A - \frac{1}{2}I & 0 \end{bmatrix}$ to find an eigenvector corresponding to $\lambda = \frac{1}{2}$:

$$\begin{bmatrix} \frac{3}{2} & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{7}{2} & 0 \end{bmatrix} \overset{2R_1}{\underset{2R_3}{\sim}} \begin{bmatrix} 3 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 7 & 0 \end{bmatrix}$$

$$\overset{2R_3+R_1}{\sim} \begin{bmatrix} 1 & 0 & -9 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 7 & 0 \end{bmatrix}$$

$$\overset{-2R_1+R_3}{\sim}\begin{bmatrix}1&0&-9&0\\0&0&0&0\\0&0&25&0\end{bmatrix}$$
$$\sim\begin{bmatrix}1&0&-9&0\\0&0&0&0\\0&0&1&0\end{bmatrix}$$
$$\overset{9R_3+R_1}{\sim}\begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&1&0\end{bmatrix}.$$

Here's what this is saying: if $\mathbf{v} = \begin{bmatrix}v_1 & v_2 & v_3\end{bmatrix}^T$ is an eigenvector of A corresponding to $\lambda = \frac{1}{2}$, or in other words a nontrivial solution of $(A - \frac{1}{2}I)\mathbf{v} = \mathbf{0}$, then we must have $v_1 = v_3 = 0$ and v_2 free. So one nonzero eigenvector corresponding to $\lambda = \frac{1}{2}$ is given by

$$\mathbf{v}_1 = \begin{bmatrix}0\\1\\0\end{bmatrix}.$$

Therefore $\left[\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right]$ forms a basis of the eigenspace $E_{\frac{1}{2}}$ of A.

To find an eigenvector corresponding to $\lambda = 3$, we'll now row reduce $\begin{bmatrix}A - 3I & \mathbf{0}\end{bmatrix}$:

$$\begin{bmatrix}-1&0&-1&0\\0&0&0&0\\1&0&1&0\end{bmatrix}\sim\begin{bmatrix}0&0&0&0\\0&1&0&0\\1&0&1&0\end{bmatrix}.$$

So if $\mathbf{v}$ is an eigenvector corresponding to the eigenvalue $\lambda = 3$, then we need $v_1 = -v_3$, and $v_2 = 0$. Which means that

$$\mathbf{v} = \begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix} = \begin{bmatrix}-v_3\\0\\v_3\end{bmatrix} = v_3\begin{bmatrix}-1\\0\\1\end{bmatrix}.$$

The vector $\left[\begin{smallmatrix}-1\\0\\1\end{smallmatrix}\right]$ therefore forms a basis of the eigenspace E_3 of A. □

As mentioned above, the computations in Example 2.10 can also be carried out using technology. In Octave, the computation proceeds using `eig`:

```
format short
A = [2, 0, -1; 0, 1/2, 0; 1, 0, 4];
[U,D] = eig(A)
```

Octave by design will give eigenvectors that have unit norm (i.e., magnitude of 1). In the above cell the first two columns of U are eigenvectors in E_3, while the last column is an eigenvector in $E_{\frac{1}{2}}$. In Subsection , we examine the reason why 3 shows up twice in the matrix D of eigenvalues (and why a column of U is repeated here).

As Octave is designed for numerical work, it's a little awkward to try to get it to produce symbolic answers. For symbolic mathematics, a system such as Sage (or another CAS) is more appropriate. In Sage, eigenvalues can be found like so:

`A.eigenvectors_right()` finds vectors $\mathbf{x}$ satisfying $A\mathbf{x} = \lambda\mathbf{x}$. `A.eigenvectors_left()` finds vectors $\mathbf{y}$ satisfying $\mathbf{y}^T A = \lambda\mathbf{y}^T$.

```
A = Matrix([[2, 0, -1], [0, 1/2, 0], [1, 0, 4]])
A.eigenvectors_right()
```

Octave-like output can also be produced using the `eigenmatrix_right()` method:

```
# run the previous cell first so A is defined!
D,U = A.eigenmatrix_right()
D,U
```

Algebraic and Geometric Multiplicities of Eigenvalues

There are a couple of interesting things we can note about the Example 2.10. First, each eigenvalue had *at least* one corresponding eigenvector. Second, $\lambda = 3$ was basically a "repeated" eigenvalue, since it showed up as a double root in the characteristic equation $\det(A - \lambda I) = 0$. This leads us to some terminology.

Definition 2.11 Algebraic and Geometric Multiplicity. Let A be an $n \times n$ matrix and let λ be an eigenvalue of A. The **algebraic multiplicity** of λ is defined to be the multiplicity of λ as a root of the characteristic equation. The **geometric multiplicity** of λ is defined to be the number of linearly independent eigenvectors corresponding to λ. Equivalently, the geometric multiplicity is exactly the dimension of the eigenspace corresponding to λ: $\dim E_\lambda$. ♢

In Example 2.10, $\lambda = 3$ had an algebraic multiplicity of 2. This could also be seen in the Octave and Sage results following the example, since 3 showed up twice in the diagonal matrix of eigenvalues. This was also given after using `eigenvectors_right()` immediately after the given eigenvector for $\lambda = 3$. On the other hand, the geometric multiplicity of $\lambda = 3$ was $\dim E_3 = 1$. This was represented in the repeated eigenvector in U in the Octave result and the single nonzero vector corresponding to $\lambda = 3$ in the Sage result. In this case, we say that A is *defective*.

Definition 2.12 Defective Matrices. Let A be a square matrix. If the sum of the geometric multiplicities of the eigenvalues of A is less than the sum of algebraic multiplicities of eigenvalues of A, then we say that A is **defective**. ♢

The following result gives some basic estimates for algebraic and geometric multiplicities of eigenvalues.

Theorem 2.13 Bounds on Algebraic and Geometric Multiplicities. *If A is an $n \times n$ matrix, then A has exactly n (not necessarily distinct!) eigenvalues, counting multiplicities. Furthermore, if λ is an eigenvalue of A, then the geometric multiplicity is always less than or equal to the algebraic multiplicity.*

From Theorem 2.13, an $n \times n$ square matrix A is defective if and only if it has at most $n - 1$ linearly independent eigenvectors. Equivalently, the eigenvectors of A are not enough to form a basis (also known as an *eigenbasis*) of $\mathbb{R}^n$. This leads to difficulties in computations involving A.

Definition 2.14 Eigenbases. Let A be an $n \times n$ square matrix that is not defective. A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is called an **eigenbasis** for $\mathbb{R}^n$ if each vector in the set is an eigenvector of A and if the set also forms a basis of $\mathbb{R}^n$. ♢

Example 2.15 Multiplicities and Eigenspaces. Graph the eigenspaces of the matrix

$$A = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}.$$

What are the algebraic and geometric multiplicities of each eigenvalue?

Solution. Since A is triangular, its eigenvalues are 0 and 1. We also know at this point that A can't be defective. In particular, we can form a basis of $\mathbb{R}^2$ entirely from eigenvectors of A. To actually find the eigenvectors, we need to find the corresponding eigenspaces $E_0 = \operatorname{null}(A - 0I) = \operatorname{null} A$ and $E_1 = \operatorname{null}(A - I)$.

By inspection,

$$\operatorname{null} A = \operatorname{span} \begin{bmatrix} -3 \\ 1 \end{bmatrix}.$$

The eigenspace associated with 1 is $\operatorname{null}(A - I)$. Since

$$A - I - \begin{bmatrix} 0 & 3 \\ 0 & -1 \end{bmatrix}$$

we see that

$$\operatorname{null}(A - I) = \operatorname{span} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

So the graph of the eigenspaces is given by

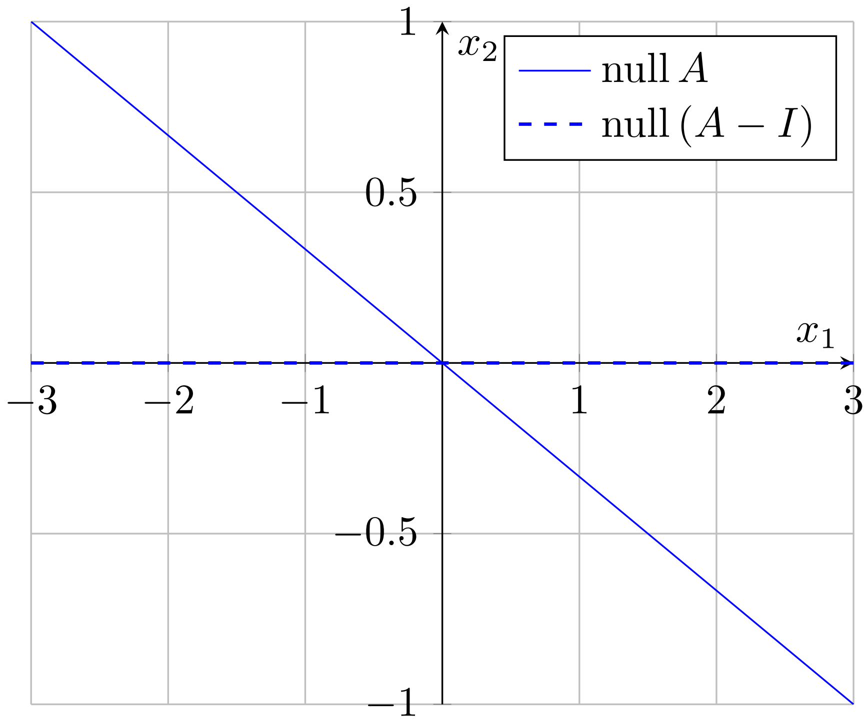

Figure 2.16 Eigenspaces of A

A corresponding eigenbasis from A would be the set $\left\{ \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$. □

The importance of an eigenbasis is demonstrated in the next example.

Example 2.17 Matrix Multiplication and Eigenvectors. Let

$$\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\} = \left\{ \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

and let A be as in the previous example. Suppose that $\mathbf{v}$ is a vector in $\mathbb{R}^2$ such that

$$\mathbf{v} = 4\mathbf{b}_1 + 3\mathbf{b}_2.$$

Compute $A\mathbf{v}$.

Solution. This computation will be quite easy. Since $\mathbf{v} = 4\mathbf{b}_1 + 3\mathbf{b}_2$, we have

$$A\mathbf{v} = 4A\mathbf{b}_1 + 3A\mathbf{b}_2 = 3\mathbf{b}_2 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

□

2.2 Eigenvalue Problems

Many important problems in mathematics and its applications reduce to statements of the form $A\mathbf{x} = \lambda\mathbf{x}$. Naturally, eigenvalues and eigenvectors are useful tools for tackling these problems. As a first example, we'll again consider a Markov process.

Example 2.18 Long-term Behavior of Markov Processes. Recall that a Markov process describes the evolution of one state $\mathbf{x}_n$ into a future state $\mathbf{x}_{n+1}$ using the matrix equation $A\mathbf{x}_n = \mathbf{x}_{n+1}$. In such a process, the matrix A is a square matrix with non-negative entries whose columns sum to 1. Starting from an initial state $\mathbf{x}_0$, we are often interested in whether the long-term evolution approaches a specific state vector $\mathbf{x}$. In symbols, we want to determine if $A^n\mathbf{x}_0 \to \mathbf{x}$ as $n \to \infty$.

For such a vector, we should have $A\mathbf{x} = \lim_{n\to\infty} A(A^n\mathbf{x}) = \mathbf{x}$. In other words, $\mathbf{x}$ is an *eigenvector of* A *with eigenvalue* 1. We call this vector a **steady-state vector** of the Markov process.

Now let's suppose we model the weather with a Markov process with transition matrix

$$A = \begin{bmatrix} 0.33 & 0.25 & 0.40 \\ 0.52 & 0.42 & 0.40 \\ 0.15 & 0.33 & 0.20 \end{bmatrix},$$

corresponding states S, C and R (i.e., "sunny", "cloudy" and "rainy") and we use an initial state vector of $\mathbf{x}_0 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$. To figure out the long-term probability that it will be a cloudy day, we can try computing $A^n\mathbf{x}_0$ for large values of n. See the Octave cell below. If we do so, it appears that the long-term probability of a cloudy day settles in around 44.6%.

We can make this analysis more precise by looking for the steady state vector using `eig`. If we take this approach, then we see that A has 1 as an eigenvalue and corresponding eigenvector $\begin{bmatrix} 0.52 & 0.74 & 0.41 \end{bmatrix}^T$. This is *not* a state vector since the values do not add to 1 and therefore can't be probabilities. However, we can convert this into a state vector by dividing each entry by the sum $0.52 + 0.74 + 0.41 = 1.678$ which in turn gives the eigenvector

$$\mathbf{x} = \begin{bmatrix} 0.3113 \\ 0.4463 \\ 0.2425 \end{bmatrix}$$

confirming our earlier guess. We can also see the long-term probabilities of sunny and rainy days from this steady-state vector as well. □

```
% Code for Markov example
format short
A = [0.33, 0.25, 0.40; 0.52, 0.42, 0.40; 0.15, 0.33, 0.20]
[U,D] = eig(A)
```

Example 2.19 Singular Values and the Condition Number. In numerical linear algebra, the **condition number** of an invertible matrix A gives an estimate of how solutions of $A\mathbf{x} = \mathbf{b}$ can change in the presence of error. More precisely, the condition number measures the response of the solution $\mathbf{x}$ if $\mathbf{b}$ is perturbed by an error term. If the condition number is small then we expect a small change in $\mathbf{x}$ if the error in $\mathbf{b}$ is small. If the condition is large, however, small changes in $\mathbf{b}$ can have significant effects on $\mathbf{x}$.

The condition number itself is denoted $\kappa(A)$. If we let $A\mathbf{x} = \mathbf{b}$ denote the unperturbed system and $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ denote the corresponding perturbed system, then the relative error between $\hat{\mathbf{x}}$ and $\mathbf{x}$ is at most equal to $\kappa(A)$ times the relative error between $\hat{\mathbf{b}}$ and $\mathbf{b}$.

The condition number $\kappa(A)$ itself can be calculated from the eigenvalues of $A^T A$ as follows:

$$\kappa(A) = \frac{\sqrt{\lambda_{\text{max}}(A^T A)}}{\sqrt{\lambda_{\text{min}}(A^T A)}}.$$

Using this, find the $\kappa(A)$ for

$$A = \begin{bmatrix} 1 & 3 & -2 \\ 0 & 2 & 4 \\ -3 & 2 & 3 \end{bmatrix}$$

□

```
% code cell for condition number example
```

2.3 Orthogonal Transformations

Two fundamental concepts in vector geometry are the *magnitude* and the *inner product*. In $\mathbb{R}^n$ these topics are related as given by the following definition.

Definition 2.20 Inner Product and Magnitude. Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^n$. The **inner product** of $\mathbf{x}$ and $\mathbf{y}$ is the real scalar $\langle \mathbf{x}, \mathbf{y} \rangle$ given by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T \mathbf{x}.$$

The **magnitude** of $\mathbf{x}$ is the nonnegative real number $\|\mathbf{x}\|$ given by

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

◊

If you've taken Calculus III, then some of the following properties will be familiar with their analogues for the dot product (see here[2]).

Theorem 2.21 Properties of the Inner Product. *Let* $\mathbf{x}, \mathbf{y}$ *and* $\mathbf{z}$ *be vectors in* $\mathbb{R}^n$ *and let* α *and* β *be real scalars. Then the inner product satisfies the following properties:*

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
2. $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$
3. $\langle \alpha\mathbf{x}, \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$ *and* $\langle \mathbf{x}, \beta\mathbf{y} \rangle = \beta \langle \mathbf{x}, \mathbf{y} \rangle$
4. $\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\|\|\mathbf{y}\| \cos\theta$ *where* θ *denotes the angle between* $\mathbf{x}$ *and* $\mathbf{y}$ *where* $0 \leq \theta \leq \pi$.

[2] j-oldroyd.github.io/wvwc-calculus/section-the-dot-product.html

5. $\langle A\mathbf{x}, \mathbf{y}\rangle = \left\langle \mathbf{x}, A^T\mathbf{y}\right\rangle$ *for any* $n \times n$ *matrix* A.

The last property in Theorem 2.21 is particularly important and can be taken as the definition of A^T. We also have the following very important inequalities involving inner products and magnitudes.

Theorem 2.22 Cauchy-Schwarz Inequality. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. *Then*

$$|\langle \mathbf{x}, \mathbf{y}\rangle| \leq \|\mathbf{x}\|\|\mathbf{y}\|.$$

Theorem 2.23 Triangle Inequality. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. *Then*

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

As the magnitude and inner product are both fundamental concepts in vector geometry, any transformation (i.e., any matrix) that preserves both of these quantities are particularly useful to work with. Such matrices are called *orthogonal transformations*.

Definition 2.24 Orthogonal Transformation. Let U be a real $n \times n$ matrix. We say that U is **orthogonal** if

$$UU^T = U^TU = I.$$

The set of all orthogonal $n \times n$ matrices is denoted by $O(n)$. ◊

Geometrically, the action of an orthogonal matrix on vectors preserves angles between vectors as measured by the inner product.

Theorem 2.25 Orthogonal Transformations and Inner Products. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *and let* U *be an* $n \times n$ *orthogonal matrix. Then*

$$\langle \mathbf{x}, \mathbf{y}\rangle = \langle U\mathbf{x}, U\mathbf{y}\rangle\,.$$

Furthermore, $\|\mathbf{x}\| = \|U\mathbf{x}\|$.

As orthogonal matrices are invertible, it follows that their columns form a basis. Such as basis has some very useful characteristics. To be precise, let $U = \left[\, \mathbf{u}_1 \cdots \mathbf{u}_n \,\right]$ denote an $n \times n$ orthogonal matrix. Then the fact that $U^TU = I$ implies that

$$\langle \mathbf{u}_i, \mathbf{u}_j\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

This leads to the following definition.

Definition 2.26 Orthonormal Basis. Let $\{\mathbf{u}_i\}_{i=1}^n$ denote a collection of vectors in $\mathbb{R}^n$. This collection is an **orthonormal basis (ONB)** if

$$\langle \mathbf{u}_i, \mathbf{u}_j\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

◊

Geometrically, an ONB in $\mathbb{R}^n$ is a collection of n orthogonal unit vectors. These can be viewed as a generalization of the typical coordinate axes.

2.4 Diagonalization of Matrices

In this section we consider bases of $\mathbb{R}^n$ that are associated with eigenvectors a square matrix A known as eigenbases (see Definition 2.14). The benefit to looking at such a basis instead of using the standard basis of $\mathbb{R}^n$ is that the eigenbasis will make products involving A much simpler through a process known as *diagonalization*.

Eigenbases

Recall that a basis of $\mathbb{R}^n$ is a linearly independent collection of n vectors in $\mathbb{R}^n$ (see also Definition 1.2). The defining characteristic of a basis is this: if $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis of $\mathbb{R}^n$ and if $\mathbf{x} \in \mathbb{R}^n$, then there exists a unique set of scalars $c_1, \ldots, c_n$ such that

$$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{b}_i.$$

This makes it possible to use the basis as a coordinate system in $\mathbb{R}^n$. Therefore we may view an eigenbasis (Definition 2.14) of a matrix A as a particular coordinate system that is well-suited to calculations involving A, an idea which we make precise below.

Example 2.27 Using an Eigenbasis to Compute a Matrix Product. Let

$$A = \begin{bmatrix} -2 & 3 \\ -4 & 5 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} -2 \\ 10 \end{bmatrix}.$$

Given that

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \mathbf{v}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

are eigenvectors of A with corresponding eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$, find $A^{100}\mathbf{b}$.

Solution. First, note that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis of $\mathbb{R}^2$. Therefore it's an eigenbasis since each vector is an eigenvector of A.

Bases in $\mathbb{R}^2$. One way to see that $\{\mathbf{v}_1, \mathbf{v}_2\}$ must be a basis of $\mathbb{R}^2$ is to observe that $\mathbf{v}_2$ is not a scalar multiple of $\mathbf{v}_1$ and so the two vectors are linearly independent. Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a set of two linearly independent vectors in the two-dimensional vector space $\mathbb{R}^2$, and so this set must be a basis of $\mathbb{R}^2$.

Since $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis of $\mathbb{R}^2$ then there exist scalars c_1, c_2 such that $\mathbf{b} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. We can find these scalars by row reduction of the augmented matrix $\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{b} \end{bmatrix}$. This reduces to

$$\begin{bmatrix} 1 & 0 & -38 \\ 0 & 1 & 12 \end{bmatrix},$$

and so $c_1 = -38, c_2 = 12$ and

$$\mathbf{b} = -38\mathbf{v}_1 + 12\mathbf{v}_2.$$

Now that we've written $\mathbf{b}$ in terms of the eigenbasis $\{\mathbf{v}_1, \mathbf{v}_2\}$, the computation of $A^{100}\mathbf{b}$ becomes almost trivial:

$$\begin{aligned} A^{100}\mathbf{b} &= A^{100}(-38\mathbf{v}_1 + 12\mathbf{v}_2) \\ &= -38A^{100}\mathbf{v}_1 + 12A^{100}\mathbf{v}_2 \\ &= -38\mathbf{v}_1 + 12 \cdot 2^{100}\mathbf{v}_2 \\ &= \begin{bmatrix} -38 + 36(2^{100}) \\ -38 + 48(2^{100}) \end{bmatrix}. \end{aligned}$$

The second to last line of this computation makes use of the fact that

$$A^n\mathbf{x} = \lambda^n\mathbf{x}$$

for any eigenvector $\mathbf{x}$ of A with eigenvalue λ.

□

```
format short
A = [1, 3, -2; 1, 4, 10]
rref(A) # reduces to find c_1, c_2
```

Example 2.27 shows that the existence of an eigenbasis can greatly simplify certain computations. Unfortunately, not every matrix has a corresponding eigenbasis (see Definition 2.12). However, the following theorem gives a simple condition that can be used to guarantee the existence of an eigenbasis.

Theorem 2.28 Distinct Eigenvalues and Eigenbases. *Let A be an $n \times n$ matrix and suppose that A has n distinct eigenvalues (equivalently, no eigenvalue is repeated). Then A has an eigenbasis.*

Proof. The proof of this statement follows from the fact that eigenvectors corresponding to distinct eigenvalues must be linearly independent, so we'll prove this first. So let $\{\lambda_i\}_{i=1}^n$ denote the eigenvalues of A and let $\mathbf{x}_i$ denote an eigenvector of A corresponding to λ_i. We'll show that $\{\mathbf{x}_i\}_{i=1}^n$ is a linearly independent set. As this will then be a set of n linearly independent vectors in $\mathbb{R}^n$, this is enough to show that it's a basis as well.

Suppose that we have scalars $\{c_i\}$ such that $\sum_{i=1}^n c_i\mathbf{x}_i = \mathbf{0}$. We need to show that $c_1 = \ldots = c_n = 0$. Now, since each $\mathbf{x}_i$ is an eigenvector of A with eigenvalue λ_i, it follows that

$$A\left(\sum_{i=1}^n c_i\mathbf{x}_i\right) = \sum_{i=1}^n c_iA\mathbf{x}_i = \sum_{i=1}^n c_i\lambda_i\mathbf{x}_i.$$

Since $A\mathbf{0} = \mathbf{0}$ as well, we have

$$\sum_{i=1}^n c_i\lambda_i\mathbf{x}_i = \mathbf{0}.$$

We can also multiply the original equation $\sum_{i=1}^n c_i\mathbf{x}_i = \mathbf{0}$ by λ_1 to get

$$\sum_{i=1}^n \lambda_1 c_i\mathbf{x}_i = \mathbf{0}.$$

Subtracting the previous two equations allows us to write

$$\mathbf{0} = \sum_{i=1}^n c_i(\lambda_i - \lambda_1)\mathbf{x}_i = \sum_{i=2}^n c_i(\lambda_i - \lambda_1)\mathbf{x}_i = \sum_{i=2}^n d_i\mathbf{x}_i$$

where $d_i = c_i(\lambda_i - \lambda_1)$. Now we can repeat the above process and write

$$\sum_{i=2}^n d_i\lambda_i\mathbf{x}_i = \mathbf{0} = \sum_{i=2}^n d_i\lambda_2\mathbf{x}_i$$

which gives (after subtracting)

$$\mathbf{0} = \sum_{i=2}^n d_i(\lambda_i - \lambda_2)\mathbf{x}_i = \sum_{i=3}^n d_i(\lambda_i - \lambda_2)\mathbf{x}_i = \sum_{i=3}^n e_i\mathbf{x}_i$$

where $e_i = (\lambda_i - \lambda_2)d_i = (\lambda_i - \lambda_2)(\lambda_i - \lambda_1)c_i$. Continuing this process, we are left with the equation

$$\mathbf{0} = (\lambda_n - \lambda_{n-1})(\lambda_n - \lambda_{n-2}) \cdots (\lambda_n - \lambda_1)c_n\mathbf{x}_n,$$

which forces $c_n = 0$.

Now we are left with

$$\sum_{i=1}^{n-1} c_i\mathbf{x}_i = \mathbf{0}$$

since we can safely disregard $c_n\mathbf{x}_n$. But there's nothing stopping us from applying the previous trick to this new sum, which will eventually show that $c_{n-1} = 0$, and then $c_{n-2} = 0$, and so on. Therefore

$$c_1 = \ldots = c_n = 0$$

and the set $\{\mathbf{x}_i\}_{i=1}^n$ must be linearly independent, which was what we needed to prove. ∎

Diagonalization

Now we'll take a closer look at just what we did in Example 2.27 to compute $A^{100}\mathbf{b}$ (or, more simply, $A\mathbf{b}$). First, we found c_1, c_2 such that $\mathbf{b} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. If we let $P = [\,\mathbf{v}_1\ \mathbf{v}_2\,]$ then this is equivalent to solving the matrix equation

$$P\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \mathbf{b} \implies \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = P^{-1}\mathbf{b}.$$

We therefore view $P^{-1}\mathbf{b}$ as the coordinates of $\mathbf{b}$ with respect to the eigenbasis $\{\mathbf{v}_1, \mathbf{v}_2\}$. Once we had the coordinates of $\mathbf{b}$ with respect to the eigenbasis, then finding $A\mathbf{b}$ amounted to multiplying c_1 and c_2 by λ_1 and λ_2 respectively. In matrix notation, this is equivalent to computing

$$D\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \text{ where } D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Finally, we used the weights $\lambda_1 c_1, \lambda_2 c_2$ to reconstruct $A\mathbf{b}$ from $\mathbf{v}_1, \mathbf{v}_2$:

$$A\mathbf{b} = \lambda_1 c_1\mathbf{v}_1 + \lambda_2 c_2\mathbf{v}_2.$$

Therefore

$$\begin{aligned} A\mathbf{b} &= \lambda_1 c_1\mathbf{v}_1 + \lambda_2 c_2\mathbf{v}_2 \\ &= P\begin{bmatrix} \lambda_1 c_1 \\ \lambda_2 c_2 \end{bmatrix} \\ &= PD\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\ &= PDP^{-1}\mathbf{b}. \end{aligned}$$

Since this equation is true for any $\mathbf{b} \in \mathbb{R}^2$, it follows that $A = PDP^{-1}$.

The process outlined above and in Example 2.27 is known as *diagonalization*. This is only possible when A has an eigenbasis to work with, but can lead to vastly more efficient computations involving A by making use of the formula

$$A^n = PD^nP^{-1}$$

since raising diagonal matrices to a power is much simpler than raising general matrices to a power. Finding $A^{100}\mathbf{b}$ in Example 2.27 was equivalent to the following computations:

$$\begin{aligned} A^{100}\mathbf{b} &= PD^{100}P^{-1}\mathbf{b} \\ &= \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1^{100} & 0 \\ 0 & 2^{100} \end{bmatrix} \begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 10 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2^{100} \end{bmatrix} \begin{bmatrix} -38 \\ 12 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -38 \\ 12(2^{100}) \end{bmatrix} \\ &= \begin{bmatrix} -38 + 36(2^{100}) \\ -38 + 48(2^{100}) \end{bmatrix}. \end{aligned}$$

Definition 2.29 Diagonalization. A matrix A is **diagonalizable** if there exists a matrix P and a diagonal matrix D such that

$$A = PDP^{-1}.$$

◊

As mentioned above, a matrix A is diagonalizable if and only if A has an eigenbasis.

Theorem 2.30 Diagonalization and Eigenbases. *Let A be an $n \times n$ matrix. Then A is diagonalizable if and only if A has a corresponding eigenbasis $\{\mathbf{v}_i\}_{i=1}^n$.*

Proof. First, assume that A is diagonalizble. Then there exist matrices P and D, say

$$P = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix} \text{ and } D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

such that P is invertible and $A = PDP^{-1}$.

We want to show that A must have an eigenbasis. We'll do this by showing that each column $\mathbf{v}_i$ of P must be an eigenvector of A with eigenvalue λ_i. Now, since $\mathbf{v}_i = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 1\mathbf{v}_i + \cdots + 0\mathbf{v}_n$ it follows that $P^{-1}\mathbf{v}_i$ must be the vector with a single 1 in the i^{th} entry and 0s elsewhere. Therefore

$$\begin{aligned} A\mathbf{v}_i &= PDP^{-1}\mathbf{v}_i \\ &= PD \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \\ &= P \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \lambda_i \\ \vdots \\ 0 \end{bmatrix} \end{aligned}$$

$$= \lambda_i \mathbf{v}_i$$

and so $\mathbf{v}_i$ must be an eigenvector of A with eigenvalue λ_i for each i from 1 to n. Since each column of P is an eigenvector of A and since P is invertible, it follows that the columns must be a basis and, hence, an eigenbasis for A.

Now we prove the reverse direction. So assume that A has an eigenbasis $\{\mathbf{v}_i\}_{i=1}^n$ with corresponding eigenvalues $\{\lambda_i\}_{i=1}^n$. We will show that A is diagonalized by

$$P = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix} \text{ and } D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

Let $\mathbf{x} \in \mathbb{R}^n$. Then we can find the coordinates of $\mathbf{x}$ with respect to the eigenbasis $\{\mathbf{v}_i\}$ by computing $P^{-1}\mathbf{x}$.

Note that P is invertible since its columns form a basis!

It follows that applying A to $\mathbf{x}$ is equivalent to applying PD to $P^{-1}\mathbf{x}$: $DP^{-1}\mathbf{x}$ will multiply the coordinates of $\mathbf{x}$ with respect to the eigenbasis by the corresponding eigenvalues, and $PDP^{-1}\mathbf{x}$ reconstructs $A\mathbf{x}$ using the weights $DP^{-1}\mathbf{x}$ to form a linear combination of the columns of P. Therefore $A\mathbf{x} = PDP^{-1}\mathbf{x}$ and so $A = PDP^{-1}$. ■

Example 2.31 Diagonalizing a Matrix. Find matrices P and D (if possible) that diagonalize

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

Solution. By Theorem 2.30, A is diagonalizble if and only if A has an eigenbasis. Using Octave we quickly see that the eigenvalues of A are $\lambda_1 = 0, \lambda_2 = \lambda_3 = 3$. Now we need to $[\,A{-}0I\ \mathbf{0}\,]$ and $[\,A{-}3I\ \mathbf{0}\,]$:

$$\begin{bmatrix} A & \mathbf{0} \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} A - 3I & \mathbf{0} \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Therefore

$$\operatorname{null}(A) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} \text{ and } \operatorname{null}(A - 3I) = \operatorname{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Now we have everything we need to diagonalize A. Define

$$P = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \text{ and } D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Then $A = PDP^{-1}$. □

```
# code cell to use for previous example
A = [2, -1, -1; -1, 2, -1; -1, -1, 2]
eig(A)
```

Diagonalizations of Symmetric and Hermitian Matrices

Symmetric matrices have particularly nice diagonalizations. First, their eigenvalues must be limited to real numbers.

Theorem 2.32 Eigenvalues of Symmetric Matrices. *Let A be a symmetric matrix with real entries and let λ be an eigenvalue of A. Then λ must be a real number.*

The proof of Theorem 2.32 is relatively simple but requires us to expand our terminology and notation a bit. First, we redefine the inner product so that it also applies to complex vectors in $\mathbb{C}^n$.

Definition 2.33 Complex Inner Product. Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. The **(complex) inner product** of $\mathbf{x}$ and $\mathbf{y}$ is the (complex) scalar

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x}$$

where $\mathbf{y}^*$ denotes the **conjugate transpose** of $\mathbf{y}$. $\Diamond$

Now we expand our definition of symmetric matrices to include the complex case as well.

Definition 2.34 Hermitian Matrices. Let A denote a square matrix. Then A is **Hermitian** if $A = A^*$. $\Diamond$

Definition 2.34 generalizes the definition of a real symmetric matrix since $A^* = A^T$ if A only has real entries. The conjugate transpose and inner product also share many properties, including the following.

Proposition 2.35 Conjugate Transpose and Inner Product. *Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ and let A be a square matrix with complex entries. Then*

$$\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A^*\mathbf{y} \rangle .$$

Now we can prove that real symmetric matrices, and more generally Hermitian matrices, always have real eigenvalues.

Theorem 2.36 Eigenvalues of a Hermitian Matrix. *Let A denote a Hermitian matrix. Then A has real eigenvalues.*

Proof. Let $\mathbf{x}$ denote an eigenvector of A with eigenvalue λ. Then

$$\begin{aligned} \lambda \langle \mathbf{x}, \mathbf{x} \rangle &= \langle \lambda\mathbf{x}, \mathbf{x} \rangle \\ &= \langle A\mathbf{x}, \mathbf{x} \rangle \\ &= \langle \mathbf{x}, A^*\mathbf{x} \rangle \\ &= \langle \mathbf{x}, A\mathbf{x} \rangle \\ &= \langle \mathbf{x}, \lambda\mathbf{x} \rangle \\ &= \overline{\lambda} \langle \mathbf{x}, \mathbf{x} \rangle \end{aligned}$$

Since $\langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|^2 \neq 0$, it follows that $\lambda = \overline{\lambda}$. Therefore $\lambda \in \mathbb{R}$. ■

The eigenvectors of a Hermitian matrix also have nice geometric properties.

Theorem 2.37 Eigenvectors of a Hermitian Matrix. *Let A be a Hermitian matrix. Suppose that $\mathbf{x}$ and $\mathbf{y}$ are eigenvectors for two distinct eigenvalues of A, say λ_i and λ_j. Then $\mathbf{x}$ and $\mathbf{y}$ are orthogonal.*

Proof. We need to show that $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. Now,

$$\lambda_i \langle \mathbf{x}, \mathbf{y} \rangle = \langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle = \lambda_j \langle \mathbf{x}, \mathbf{y} \rangle$$

or just $(\lambda_i - \lambda_j) \langle \mathbf{x}, \mathbf{y} \rangle = 0$. Since $\lambda_i \neq \lambda_j$, it follows that $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. ■

Theorem 2.37 leads to an extremely useful fact about real symmetric matrices: they can always be *orthogonally diagonalized*. This means that we can choose an eigenbasis that is also an orthonormal basis. As an example, consider the eigenbasis found in Example 2.31, replicated here as columns of the matrix P:

$$P = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Note that the first column is orthogonal to the other two which is guaranteed by Theorem 2.37 since these vectors correspond to different eigenvalues.

Although the last two columns are not orthogonal, they can be *orthogonalized*. One way of doing so is to replace them with the vectors

$$\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

Both vectors are still eigenvectors with eigenvalue 0 and the two of them, together with $\left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right]$, still form an eigenbasis of $\mathbb{R}^3$. However, this new basis is orthogonal.

If we go one step further and normalize the vectors in this eigenbasis we then get the orthonormal eigenbasis

$$\left\{ \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \sqrt{\frac{2}{3}} \end{bmatrix} \right\}.$$

This provides the orthogonal diagonalization

$$U = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \sqrt{\frac{2}{3}} \end{bmatrix} \text{ and } D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

which gives

$$A = UDU^{-1} = UDU^T.$$

Such a diagonalization is always possible for real symmetric and Hermitian matrices, a result known as the *Spectral Theorem*.

Theorem 2.38 Spectral Theorem. *Let A be a (real) symmetric matrix. Then there exists an orthogonal matrix U and a diagonal matrix D such that*

$$A = UDU^T.$$

Equivalently, if $\{\mathbf{u}_i\}_{i=1}^n$ is an orthonormal eigenbasis of A with corresponding eigenvalues $\{\lambda_i\}_{i=1}^n$, then

$$A = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T.$$

Example 2.39 A Signal Processing Application. In the mathematics of signal processing, signals are often represented as particular vectors in $\mathbb{R}^n$. The problem of signal transmission then reduces to sending a list of numbers $c_1, \ldots, c_m$ such that the receiver can use these to reconstruct a signal $\mathbf{x}$. Such a scheme can be implemented by choosing an appropriate $n \times m$ matrix

$F = [\, \mathbf{f}_1 \dots \mathbf{f}_m \,]$ and then computing and transmitting the coordinates of

$$F^T\mathbf{x} = \left[\langle \mathbf{x}, \mathbf{f}_i \rangle\right]_{1 \le i \le m}.$$

The important properties of the columns of F can then be encoded in the *Gram matrix* $G = F^T F = \left[\langle \mathbf{f}_i, \mathbf{f}_j \rangle\right]$. Find a 3×4 matrix F for which that Gram matrix is

$$G = \begin{bmatrix} 1 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & 1 & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 1 \end{bmatrix}.$$

Solution. This problem can be solved by diagonalizing G: if $G = UDU^T$ then we can define F by using the rows of $U\sqrt{D}$.

If D is diagonal with nonnegative entries, then we can define $\sqrt{D}$ to be the matrix obtained by taking the square root of each entry of D. It's not obvious, but it turns out that Gram matrices always have nonnegative eigenvalues. If we diagonalize G and find it has negative eigenvalues, then it cannot be a Gram matrix.

So we'll diagonalize G with the help of the Octave cell below. Doing so, we see that the eigenvalues of G are in fact nonnegative and so $U\sqrt{D}$ will only contain real values. Furthermore, U only has 3 nonzero columns since G only has three nonzero eigenvalues. By removing the zero column from $U\sqrt{D}$, we obtain a 4×3 matrix which we define to be F^T. Then $F^T F = G$. □

```
# diagonalizing a Gram matrix
format short
G = (4/3)*eye(4) - (1/3)*ones(4) # tricky way to define G
```

Symmetric matrices are also useful in analyzing *quadratic forms*, which are expressions of the form

$$\sum_{i,j} c_{ij} x_i x_j$$

where the x_i are variables and the c_{ij} are the coefficients. Such an expression can be rewritten as $\mathbf{x}^T A \mathbf{x}$ where A is a symmetric matrix determined from the coefficients.

Example 2.40 Analyzing a Quadratic Form. Describe the curve given by

$$9x^2 + 6xy + y^2 = 10.$$

Solution. The left hand side is a quadratic form with variables $x_1 = x, x_2 = y$ and coefficients

$$c_{11} = 9, c_{12} = c_{21} = \frac{6}{2} = 3 \text{ and } c_{22} = 1.$$

Therefore we can write $9x^2 + 6xy + y^2 = \mathbf{x}^T A \mathbf{x}$ where

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \text{ and } A = \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}.$$

To help us describe the curve we will "disentangle" the variables x and y by diagonalizing A.

Since A is symmetric, we know that A can be orthogonally diagonalized.

One such diagonalization is given by $A = UDU^T$ where

$$U = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix} \text{ and } D = \begin{bmatrix} 0 & 0 \\ 0 & 10 \end{bmatrix}.$$

The quadratic form $\mathbf{x}^T A\mathbf{x}$ then becomes

$$\mathbf{x}^T UDU^T\mathbf{x} = (U^T\mathbf{x})^T D(U^T\mathbf{x}).$$

Now we define $\mathbf{y} = U^T\mathbf{x}$ as a change-of-variables. In particular,

$$\mathbf{y} = \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{10}}x - \frac{3}{\sqrt{10}}y \\ \frac{3}{\sqrt{10}}x + \frac{1}{\sqrt{10}}y \end{bmatrix}.$$

Our quadratic form is now

$$\mathbf{y}^T D\mathbf{y} = 0X^2 + 10Y^2 = 10Y^2$$

and our original equation becomes $10Y^2 = 10$ or just $Y = \pm 1$. Therefore the original equation describes the two different lines

$$\frac{3}{\sqrt{10}}x + \frac{1}{\sqrt{10}}y = -1 \text{ and } \frac{3}{\sqrt{10}}x + \frac{1}{\sqrt{10}}y = 1.$$

□

Analytic Functions of Matrices

A function $f(x)$ is *analytic* at $x = a$ if it has a power series representation on some interval centered around a:

$$f(x) = \sum_{k=0}^{\infty} c_k(x-a)^k = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots \text{ for } x \approx a.$$

If A denotes a square matrix, and if $f(x)$ is analytic at $a = 0$, then we can try to make sense of the expression $f(A)$ by using the power series for $f(x)$:

$$f(A) = \sum_{k=0}^{\infty} c_k A^k = c_0 I + c_1 A + c_2 A^2 + \cdots,$$

assuming this sum actually exists.

Example 2.41 Exponential of a Matrix. Let

$$A = \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Find e^A.

Solution. By definition,

$$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots,$$

so we can find e^A by looking at the powers of A. In this case,

$$A^2 = \begin{bmatrix} 0 & 0 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } A^3 = A^4 = \ldots = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore

$$e^A = I + A + \frac{1}{2}A^2 = \begin{bmatrix} 1 & 2 & 5 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}.$$

This can be verified using the command `expm` in Octave as below.

You want to be careful to use `expm(A)` to compute e^A in Octave, as this actually finds the matrix exponential. If you instead use `exp(A)`, this just computes the matrix obtained by raising e to each entry of A, which in general is *not* equal to e^A.

□

```
format short
A = [0, 2, 3; 0, 0, 2; 0, 0, 0]
expm(A) # don't use exp(A)! that just exponentiates each
    entry of A
```

Example 2.42 Function of a Diagonal Matrix. Let $f(x)$ be a function analytic at 0. Let D be a diagonal matrix whose diagonal entries are within the interval of convergence of the power series for $f(x)$ centered at 0. Find $f(D)$.

□

The last example shows that if $f(x)$ is analytic at 0 then it is relatively straightforward to find $f(D)$, assuming that the diagonal entries of D are within the interval of convergence for the series representation of $f(x)$ at 0. Therefore we can find functions of diagonalizable matrices as well.

Theorem 2.43 Analytic Functions of Diagonalizable Matrices. *Let $f(x)$ be a function that is analytic at 0 and whose series representation at 0 has interval of convergence I. Let A be a diagonalizable matrix diagonalized by P and D with eigenvalues contained in I. Then*

$$f(A) = Pf(D)P^{-1}.$$

Proof. Let $f(x) = \sum_{k=0}^{\infty} c_k x^k$ and recall that $A^k = PD^kP^{-1}$. Then

$$f(A) = \sum_{k=0}^{\infty} c_k A^k = P\left(\sum_{k=0}^{\infty} c_k D^k\right)P^{-1} = Pf(D)P^{-1}.$$

■

Theorem 2.43 allows for straightforward computations of matrix exponentials of diagonalizable matrices. This is useful in differential equations when solving linear systems of ODEs (see here[3]).

Theorem 2.44 Exponential Solutions of Linear Systems of ODEs. *Let A be a (constant) square matrix and consider the first-order system $\mathbf{x}' = A\mathbf{x}$ with initial condition $\mathbf{x}(0) = \mathbf{x}_0$. Then the solution of this initial value problem is*

$$\mathbf{x}(t) = e^{At}\mathbf{x}_0.$$

Proof. The proof follows quickly from the fact that $\frac{\mathrm{d}}{\mathrm{d}t} = Ae^{At}$. Using this, we differentiate $\mathbf{x}(t) = e^{At}\mathbf{x}_0$ to get

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = Ae^{At}\mathbf{x}_0 = A\mathbf{x}.$$

Furthermore, $\mathbf{x}(0) = e^{\mathbf{0}}\mathbf{x}_0 = \mathbf{x}_0$. Therefore $\mathbf{x}(t) = e^{At}\mathbf{x}_0$ is a solution of the

[3] j-oldroyd.github.io/wvwc-differential-equations/systems-of-odes.html

initial value problem. ■

Theorems 2.43 and 2.44 allow us to find solutions of linear systems that involve diagonalizable matrices in terms of the matrix exponential.

Example 2.45 Solving a First-Order System. Solve

$$x' = 3x - 4y \text{ and } y' = -4x + 3y$$

with initial condition $x(0) = 2$ and $y(0) = -1$.

Solution. This can be solved very easily using the matrix exponential and diagonalization. If we let

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \text{ and } A = \begin{bmatrix} 3 & -4 \\ -4 & 3 \end{bmatrix}$$

then the system can be written as the matrix ODE $\mathbf{x}' = A\mathbf{x}$ with initial condition $\mathbf{x}_0 = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$. A is symmetric and can be orthogonally diagonalized by

$$U = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \text{ and } D = \begin{bmatrix} -1 & 0 \\ 0 & 7 \end{bmatrix}$$

as seen in the code cell below this example. Therefore $A = UDU^T$, $e^{At} = Ue^{Dt}U^T$ and the solution of the system must be

$$\mathbf{x} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{7t} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

□

```
format short
A = [3, -4; -4, 3]
[U,D] = eig(A)
```

Part II

Multivariable Calculus

Chapter 3

Vector Derivatives

In this chapter we review important concepts relating to vector functions and their derivatives. As in introductory calculus, the derivative may be seen as giving the rate of change of a particular quantity. However, the move to higher dimensions involved with vector functions allows for multiple interpretations of the rate of change of a vector function, and therefore several different notions of the derivative.

Appendices

Appendix A

GNU Free Documentation License

Version 1.3, 3 November 2008

0. PREAMBLE. The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS. This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

The “publisher” means any person or entity that distributes copies of the Document to the public.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING. You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY. If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS. You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties — for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS. You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS. You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS. A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION. Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION. You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE. The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING. "Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents. To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C)  YEAR  YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.3
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with… Texts." line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Index

Colophon

This book was authored in MathBook XML.