LINEAR ALGEBRA

OR

E Pluribus Unum

Dr. Ralph deLaubenfels

TABLE OF CONTENTS

Introduction: Quick history, motivation, prerequisites. page 1

Equations, rabbits and foxes. E pluribus unum will be explained.

Chapter I: Matrices and vectors. page 13

Section A: Matrices, examples and algebra. page 15

A unifying theme for the class. Much of our use of matrices is "just" terminology, but it's really *great* terminology.

Section B: Vectors. page 34

Ordered n-tuples of real numbers, terminology $\mathbf{R}^n$. Figuratively represented by an arrow, literally for $n = 2$, indicating both magnitude and direction. The arrows also provide a geometric picture of addition of vectors and multiplication of vectors by real numbers.

Section C: Difference equations. page 47

We think of matrix multiplication as a verb. More precisely, the identity of a matrix A is the relationship between a vector $\vec{x}$ and the product $A\vec{x}$. This will be of particular interest when $\vec{x}$ is a set of populations or other physical states, that undergo change at regular intervals of time: the matrix A will represent that change of state, in the sense that, if $\vec{x}$ is the state now, then $A\vec{x}$ will be the state in the next unit of time, e.g., the next year, the next day, the next hour, etc. This set-up is called a *difference equation.* This section is mostly presentation without explanation; Chapter VII will show how to construct the matrices A driving the changing states, while Chapter VIII will show how to solve the corresponding difference equations.

Chapter II: Linear systems and Gauss-Jordan elimination. page 75

Section A: Linear systems, definitions and terminology. page 77

A linear system is a bunch of linear equations. As with Chapter I, the terminology is well worth the effort to learn; it will make everything look one dimensional and will hint at future deep ideas. For pairs of equations with two variables, lines in the plane will be drawn, and the relevance of professional wrestling will be made clear.

Section B: Gauss-Jordan elimination. page 100

This is the technique you must use for solving linear systems. Getting the right answer to a linear system is not enough, and is often of less importance than other matters associated with the linear system. The techniques of this section will reappear often in future, potentially more mysterious settings, and will change unfamiliar ideas or calculations into the familiar one of solving linear systems.

Chapter III: Qualitative information about linear systems and matrices. page 129

Section A: More matrix matters. page 132

Ideas not needed for Chapter II, hence not included in Chapter I, since we wanted to get to linear systems. Diagonal matrices are of particular interest, representing an ideal that all matrices will try to emulate, especially when we solve difference equations in Chapter VIII, or (for those who have had calculus) systems of constant-coefficient differential equations in Appendix Two. The transpose of a matrix will make unexpected appearances in this section and the future.

Section B: Rank. page 145

To each matrix there is associated a nonnegative integer called the *rank* of the matrix. Very informally, rank is telling you how much information a matrix contains. When the matrix is the coefficient matrix of a linear system, this information is negative, in the sense of decreasing the number of degrees of freedom in the set of solutions of the linear system.

This section begins with homogeneous linear systems and the number of solutions $(0, 1, \text{ or } \infty)$ that a linear system can have. By counting the number of *free variables* in the set of solutions of a linear system, we can distinguish between different infinities, e.g., the infinity of a line versus the infinity of a plane.

Rank, when applied to linear systems, will tell us if a linear system is consistent (has a solution), and, if it is consistent, how many free variables are in the set of solutions, without having to solve the linear system.

Rank is a unifying, simplifying theme that will reappear often throughout this book.

Section C: Null space and range space. page 188

Both of these spaces refer to matrices, or, in the setting of Chapter VII, linear transformations. If A is an $(m \times n)$ matrix, the *null space* of A is the set of all vectors $\vec{x}$ in $\mathbf{R}^n$ such that $A\vec{x} = \vec{0}$, while the *range space* of A is the set of all $\{A\vec{x} \,|\, \vec{x} \text{ is in } \mathbf{R}^n\}$; notice that null space is contained in $\mathbf{R}^n$ and range space is contained in $\mathbf{R}^m$. This leads to the definition of a *singular* matrix, one which has nontrivial null space.

Both null space and range space can be expressed in terms of linear systems. Writing linear systems in matrix form, the null space of A is the set of all solutions of the homogeneous linear system $A\vec{x} = \vec{0}$, while the range space of A is the set of all vectors $\vec{b}$ for which $A\vec{x} = \vec{b}$ is consistent.

Null space is relevant to *uniqueness* of solutions of linear systems, while range space is relevant to *existence* of solutions.

Section D: Invertibility and inverses. page 207

The matrix form $A\vec{x} = \vec{b}$ (A is a matrix, $\vec{x}$ and $\vec{b}$ are vectors) of a linear system makes many equations look like one equation. To make this appearance complete, we'd like to solve by multiplying both sides by A^{-1}: just as $2x = 5$ implies $x = 2^{-1} \cdot 5 \equiv \frac{5}{2}$, $A\vec{x} = \vec{b}$ should imply that our solution is $A^{-1}\vec{b}$. This section attempts to make sense out of A^{-1}, the *inverse* of the matrix A.

On a more visceral level, (multiplication by) A^{-1} will undo whatever damage (multiplication by) A did.

Chapter IV: Vector spaces, construction and maintenance. page 228

Section A: Linear combinations. page 231
Here we do everything we know how to do with vectors: add, and multiply by real numbers.

Section B: Vector spaces and span. page 247
Span of a set of vectors means the set of all possible linear combinations. A *vector space* is a set of vectors that is closed under linear combinations; that is, linear combinations of vectors in the set remain in the set. The vector spaces we will focus on are spans of specified sets of vectors and null spaces of specified matrices.

Section C: Linear dependence. page 263
A set of vectors is *linearly dependent* if one vector in the set is a linear combination of the others. From the point of view of span, a linearly dependent set has unnecessary vectors; a vector could be removed without changing the span.

A matrix is singular (see Chapter IIIC) if and only if its columns are linearly dependent.

Section D: Basis. page 283
A *basis* for a vector space V is a linearly independent set that spans V. These two requirements are working at cross purposes: for span you want more vectors, for linear independence you want fewer. This section focusses on constructing bases for our favorite vector spaces, null space of a matrix and span of a set of vectors.

Section E: Dimension. page 304
As we meant to imply in the second sentence of Section D, it can be shown that, for any vector space V, any pair of bases for V have the same number of vectors. That number is called the *dimension* of V. The intuition of dimension is the number of free variables. It is also the minimum number of spanning vectors and the maximum number of linearly independent vectors.

As in Section D, we focus on dimension of the null space of a matrix and the span of a set of vectors.

Section F: Value-based equivalences. page 337
We were concerned at this point at the number of recent definitions that came in pairs, e.g., linear dependence versus linear independence. These are the easiest to misremember, so in this section we associated "goodness" or "badness" to each member of a pair. For example, we called an invertible matrix A "good" because then the linear system $A\vec{x} = \vec{b}$ has the unique solution $\vec{x} = A^{-1}\vec{b}$, for any vector $\vec{b}$.

Chapter V: Determinants. page 353

Determinants, defined for square matrices, are an interesting combination of algebra and geometry. The definition is grubbily algebraic, but many of its properties are geometric; for example, for vectors $\vec{a} \equiv (a_1, a_2)$ and $\vec{b} \equiv (b_1, b_2)$, the parallelogram formed by $\vec{a}$ and $\vec{b}$ has area equal to the absolute value of the determinant of $\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$.

The characterization of a square matrix being invertible if and only if its determinant is nonzero will be crucial in Chapter VIII.

Chapter VI: Norm and orthogonality. page 389

Section A: Norm. page 393

For a two-dimensional vector $\vec{x} \equiv (x_1, x_2)$, the *norm* of $\vec{x}$ is the length of an arrow representing $\vec{x}$, $\|\vec{x}\| \equiv \sqrt{x_1^2 + x_2^2}$. We quite naturally extend this to $\|(x_1, x_2, \ldots, x_n)\| \equiv \sqrt{x_1^2 + x_2^2 + \ldots x_n^2}$.

Section B: Orthogonality and dot product. page 397

After demonstrating, with hot sand and cool water, the importance of orthogonality (making a right angle), we show that two vectors $\vec{x} \equiv (x_1, x_2, \ldots, x_n)$ and $\vec{y} \equiv (y_1, y_2, \ldots, y_n)$ are *orthogonal* or *perpendicular*, denoted $\vec{x} \perp \vec{y}$, if and only if their *dot product* $\vec{x} \cdot \vec{y} \equiv (x_1 y_1 + x_2 y_2 + \ldots x_n y_n)$ is zero.

The *(orthogonal) projection* of a vector $\vec{x}$ onto a vector space W, denoted $P_W(\vec{x})$, is a vector in W such that $(\vec{x} - P_W(\vec{x})) \perp W$; informally, we are dropping a perpendicular onto W. We show what our intuition suggests, that $P_W(\vec{x})$ is the best approximation of $\vec{x}$ from W, that is, the point in W closest to $\vec{x}$.

For W equal to the span of a fixed vector $\vec{b}$, we get an algebraic formula for $P_W(\vec{x}) \equiv \mathrm{proj}_{\vec{b}}(\vec{x})$.

Section C: Orthogonal sets and bases. page 433

The advantages of *orthogonal* (as opposed to merely linearly independent) sets and bases are extolled.

Section D: Gram-Schmidt orthogonalization. page 464

This is a particularly desirable way of changing a linearly independent set into an orthogonal set.

Section E: Least-squares solutions. page 472

Any engineer, scientist, statistician, or philosopher has to deal with uncertainty and error. For a linear system this might mean no solutions, even though you know that there should be a solution. We take the following constructive response. If $\vec{x}_0$ is a solution of $A\vec{x} = \vec{b}$, then $\|A\vec{x}_0 - \vec{b}\| = 0$. Rather than insist on a solution, we settle for *minimizing* $\|A\vec{x} - \vec{b}\|$. Such a vector $\vec{x}$ is called a *least-squares solution*. Being a least-squares solution is shown, via orthogonal projections, to be equivalent to satisfying the *normal equations* $A^T A\vec{x} = A^T\vec{b}$. This technique is applied to fitting a model to bivariate data $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots\}$, in particular *least-squares approximating* lines, parabolas, etc.

Section F: Vector cross products. page 529

In $\mathbf{R}^3$ only, there is another way to multiply vectors besides the dot product, what is known as the *(vector) cross product*. Properties that make it worthwhile are proved and demonstrated. We mention briefly lines and planes.

Chapter VII: Linear transformations. page 557

Section A: Linear transformations and matrices. page 560

Linear transformations are functions T from $\mathbf{R}^n$ to $\mathbf{R}^m$ with very limiting properties: $T(c\vec{x}) = cT(\vec{x})$ and $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$, for real c, vectors $\vec{x}, \vec{y}$. Linear transformations correspond to matrix multiplication, in the sense that a function T is linear if and only if there is a matrix A, called the *standard matrix* for T, such that $T(\vec{x}) = A\vec{x}$, for all $\vec{x}$, where $\vec{x}$ is written as a column when multiplied by A.

Section B: Rigid motions. page 585

Standard matrices for projection and reflection, and rotations by multiples of 45 degrees, are derived. General rotation matrices are discussed in Appendix One.

Section C: More examples, including difference equations. page 613

We derive the standard matrices for the difference equations in Section IC, and others, including those that give us the Fibonacci numbers.

Chapter VIII: Eigenvalues and eigenvectors. page 637

Here we focus on ways to make matrices act like diagonal matrices. Much time is spent applying these behaviours to difference equations: among other things, the local behaviour of Section A is applied to the *Fibonacci numbers*, and the global behaviour of Section B is applied to a predator/prey (foxes and rabbits) relationship.

Section A: Eigenvalues, eigenvectors and eigenspaces. page 646

If $A\vec{x} = \lambda\vec{x}$, where A is a matrix, $\vec{x}$ is a nontrivial vector, and λ is a real number, then λ is an *eigenvalue* for A and $\vec{x}$ is an *eigenvector* for A. This is local behaviour like a diagonal matrix; in particular, $A^n\vec{x} = \lambda^n\vec{x}$, for $n = 1, 2, 3, \ldots$.

Section B: Diagonalizing. page 685

Global behaviour like a diagonal matrix is to have a matrix $A = PDP^{-1}$, for some diagonal matrix D, invertible matrix P; A is then said to be *diagonalizable*, and satisfies $A^n = PD^nP^{-1}$, for $n = 1, 2, 3, \ldots$.

Appendix One: Rotation matrices. page 747

After some trigonometry, the standard matrices for arbitrary rotations are derived.

Appendix Two: Systems of differential equations. page 755

For those readers who have seen differentiation and exponential functions, we solve systems of constant-coefficient differential equations with Chapter VIII methods, very analogously to our solutions of difference equations.

Appendix Three: Pythagorean theorem. page 765

We prove the Pythagorean theorem, for right triangles with legs of length a and b, hypotenuse of length c, by drawing two squares with sides of length $(a + b)$, one containing a square of length c, the other containing squares of lengths a and b.

Appendix Four: Angles between vectors. page 771

As in Appendix One, this requires some trigonometry. We show how the dot product can give (or define in higher dimensions) the measure of arbitrary angles between arbitrary vectors. This generalizes our use of the dot product in Chapter VI to characterize vectors with angles of measure ninety degrees between them.

Index page 780

INTRO-DUCTION:

Quick History, Motivation, Prerequisites

Historically, linear algebra begins with the following sort of problem.

Find all values of x, y, and z ("variables") so that

$$x + 2y + 3z = 4 \quad \underline{\text{and}}$$
$$5x + 6y + 7z = 8 \qquad (*)$$

Rewrite $(*)$ as

$$\begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix} \qquad (**)$$

matrix

vectors

That rewriting of (*) as (**) was first done in the Han Dynasty, very approximately 0 AD.

The multi-variable (x, y, and z) problem has been changed into a single-variable $\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right)$ problem, hence the subtitle of this book: "E Pluribus Unum," or "out of many, one" (motto of the United States).

Although problems like (*) will provide excellent motivation for many ideas, we will see vectors and matrices take on lives of their own.

Here is an example that we will feel compelled to return to, because of its emotional and ecological impact.

Denote variables r, f as follows.

r is rabbit population
f is fox population.

Because r and f interact (a euphemism for some of that emotional value), we put them together as a vector $\begin{bmatrix} r \\ f \end{bmatrix}$.

We shall see that, in many scenarios, there will be a matrix M so that

$$M \begin{bmatrix} r \\ f \end{bmatrix}_{\text{now}} = \begin{bmatrix} r \\ f \end{bmatrix}_{\text{next year}}$$

M will neatly encapsulate all the interaction between foxes and rabbits: multiplying by M moves foxes and rabbits a year into the future.

Notice that

$$M \cdot M \begin{bmatrix} r \\ f \end{bmatrix}_{\text{now}} = \begin{bmatrix} r \\ f \end{bmatrix}_{\substack{\text{two years} \\ \text{from now}}} ;$$

in general, we can move as far into the future as we like, by multiplying by M sufficiently many times.

If we can get a sufficiently coherent expression for

$$\underbrace{M \cdot M \cdots M}_{k \text{ times}} \begin{bmatrix} r \\ f \end{bmatrix},$$

for arbitrary k, our psychic powers can extend to determining if rabbits (hence foxes) ever go extinct.

Except for the appendices, the only prerequisite for this book are a good first-year high school algebra class. In case that doesn't include the Pythagorean theorem, we prove that in Appendix Three. Appendices One and Four require some trigonometry; Appendix Two some calculus.

If the appendices are included, this book covers everything in a standard sophomore-level linear algebra college class.

IMPORTANT TERMINOLOGY.

"$\equiv$" means "is defined to be"

For example, we write

"$2 + 2 = 4$"

after counting on our fingers, but

"Brutopian $\equiv$ resident of Brutopia"

is a definition.

TWO SETS OF NUMBERS:

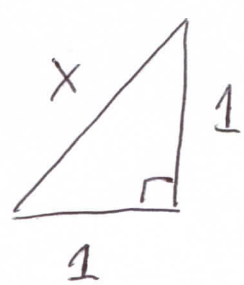
$\mathbb{N} \equiv \{\text{natural numbers}\}$

$\equiv \{1, 2, 3, 4, \ldots\}$.

Natural numbers are the result of counting. Given sufficiently many primates with the same counting goal, we can reach any natural number by counting on our fingers.

$\mathbb{R} \equiv \{ \underline{\text{real}}\ \underline{\text{numbers}} \}$

Informally, a real number is a measurement or the consequence of a measurement. For example, you could measure the legs of a right triangle to be 1:



;

then the remaining side measures

$$x = \sqrt{2},$$

by the Pythagorean theorem

(see Appendix.)

Thus $\sqrt{2}$ is a real number.

Real numbers are also called **scalars** when we want to distinguish them from matrices.

Proofs in this book will be limited to the short and intuitive. A future book will give proofs of all sorts.