Fundamentals of Linear Algebra and Optimization

CIS515, Some Slides

Jean Gallier
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
e-mail: jean@cis.upenn.edu

December 9, 2014

Contents

Chapter 1

Vector Spaces, Bases, Linear Maps

1.1 Motivations: Linear Combinations, Linear Independence and Rank

Consider the problem of solving the following system of three linear equations in the three variables $x_1, x_2, x_3 \in \mathbb{R}$:

$$
\begin{aligned}
x_1 + 2x_2 - x_3 &= 1 \\
2x_1 + x_2 + x_3 &= 2 \\
x_1 - 2x_2 - 2x_3 &= 3.
\end{aligned}
$$

One way to approach this problem is introduce some "column vectors."

Let u, v, w, and b, be the *vectors* given by

$$u = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \quad v = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} \quad w = \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

and write our linear system as

$$x_1 u + x_2 v + x_3 w = b.$$

In the above equation, we used implicitly the fact that a vector z can be multiplied by a scalar $\lambda \in \mathbb{R}$, where

$$\lambda z = \lambda \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} \lambda z_1 \\ \lambda z_2 \\ \lambda z_3 \end{pmatrix},$$

and two vectors y and and z can be added, where

$$y + z = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} y_1 + z_1 \\ y_2 + z_2 \\ y_3 + z_3 \end{pmatrix}.$$

The set of all vectors with three components is denoted by $\mathbb{R}^{3\times 1}$.

The reason for using the notation $\mathbb{R}^{3\times 1}$ rather than the more conventional notation $\mathbb{R}^3$ is that the elements of $\mathbb{R}^{3\times 1}$ are *column vectors*; they consist of three rows and a single column, which explains the superscript 3×1.

On the other hand, $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ consists of all triples of the form (x_1, x_2, x_3), with $x_1, x_2, x_3 \in \mathbb{R}$, and these are *row vectors*.

For the sake of clarity, in this introduction, we will denote the set of column vectors with n components by $\mathbb{R}^{n\times 1}$.

An expression such as

$$x_1 u + x_2 v + x_3 w$$

where u, v, w are vectors and the x_is are scalars (in $\mathbb{R}$) is called a *linear combination*.

Using this notion, the problem of solving our linear system

$$x_1 u + x_2 v + x_3 w = b$$

is equivalent to

determining whether b can be expressed as a linear combination of u, v, w.

Now, if the vectors u, v, w are *linearly independent*, which means that there is *no* triple $(x_1, x_2, x_3) \neq (0, 0, 0)$ such that

$$x_1 u + x_2 v + x_3 w = 0_3,$$

it can be shown that *every* vector in $\mathbb{R}^{3\times 1}$ can be written as a linear combination of u, v, w.

Here, 0_3 is the *zero vector*

$$0_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

It is customary to abuse notation and to write 0 instead of 0_3. This rarely causes a problem because in most cases, whether 0 denotes the scalar zero or the zero vector can be inferred from the context.

In fact, every vector $z \in \mathbb{R}^{3\times 1}$ can be written *in a unique way* as a linear combination

$$z = x_1 u + x_2 v + x_3 w.$$

Then, our equation

$$x_1 u + x_2 v + x_3 w = b$$

has a *unique solution*, and indeed, we can check that

$$\begin{aligned} x_1 &= 1.4 \\ x_2 &= -0.4 \\ x_3 &= -0.4 \end{aligned}$$

is the solution.

But then, *how do we determine that some vectors are linearly independent?*

One answer is to compute the *determinant* $\det(u, v, w)$, and to check that it is nonzero.

In our case,

$$\det(u, v, w) = \begin{vmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & -2 \end{vmatrix} = 15,$$

which confirms that u, v, w are linearly independent.

Other methods consist of computing an LU-decomposition or a QR-decomposition, or an SVD of the *matrix* consisting of the three columns u, v, w,

$$A = \begin{pmatrix} u & v & w \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & -2 \end{pmatrix}.$$

If we form the vector of unknowns

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

then our linear combination $x_1 u + x_2 v + x_3 w$ can be written in matrix form as

$$x_1u + x_2v + x_3w = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

So, our linear system is expressed by

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},$$

or more concisely as

$$Ax = b.$$

Now, what if the vectors u, v, w are *linearly dependent*?

For example, if we consider the vectors

$$u = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \qquad v = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \qquad w = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix},$$

we see that

$$u - v = w,$$

a nontrivial *linear dependence*.

It can be verified that u and v are still linearly independent.

Now, for our problem

$$x_1 u + x_2 v + x_3 w = b$$

to have a solution, it must be the case that b can be expressed as linear combination of u and v.

However, it turns out that u, v, b are linearly independent (because $\det(u, v, b) = -6$), so b cannot be expressed as a linear combination of u and v and thus, our system has *no* solution.

If we change the vector b to

$$b = \begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix},$$

then

$$b = u + v,$$

and so the system

$$x_1 u + x_2 v + x_3 w = b$$

has the solution

$$x_1 = 1, \quad x_2 = 1, \quad x_3 = 0.$$

Actually, since $w = u - v$, the above system is equivalent to

$$(x_1 + x_3)u + (x_2 - x_3)v = b,$$

and because u and v are linearly independent, the unique solution in $x_1 + x_3$ and $x_2 - x_3$ is

$$\begin{aligned} x_1 + x_3 &= 1 \\ x_2 - x_3 &= 1, \end{aligned}$$

which yields an *infinite number* of solutions parameterized by x_3, namely

$$\begin{aligned} x_1 &= 1 - x_3 \\ x_2 &= 1 + x_3. \end{aligned}$$

In summary, a 3×3 linear system may have a unique solution, no solution, or an infinite number of solutions, depending on the linear independence (and dependence) or the vectors u, v, w, b.

This situation can be generalized to any $n \times n$ system, and even to any $n \times m$ system (n equations in m variables), as we will see later.

The point of view where our linear system is expressed in matrix form as $Ax = b$ stresses the fact that the map $x \mapsto Ax$ is a *linear transformation*.

This means that

$$A(\lambda x) = \lambda(Ax)$$

for all $x \in \mathbb{R}^{3\times 1}$ and all $\lambda \in \mathbb{R}$, and that

$$A(u + v) = Au + Av,$$

for all $u, v \in \mathbb{R}^{3\times 1}$.

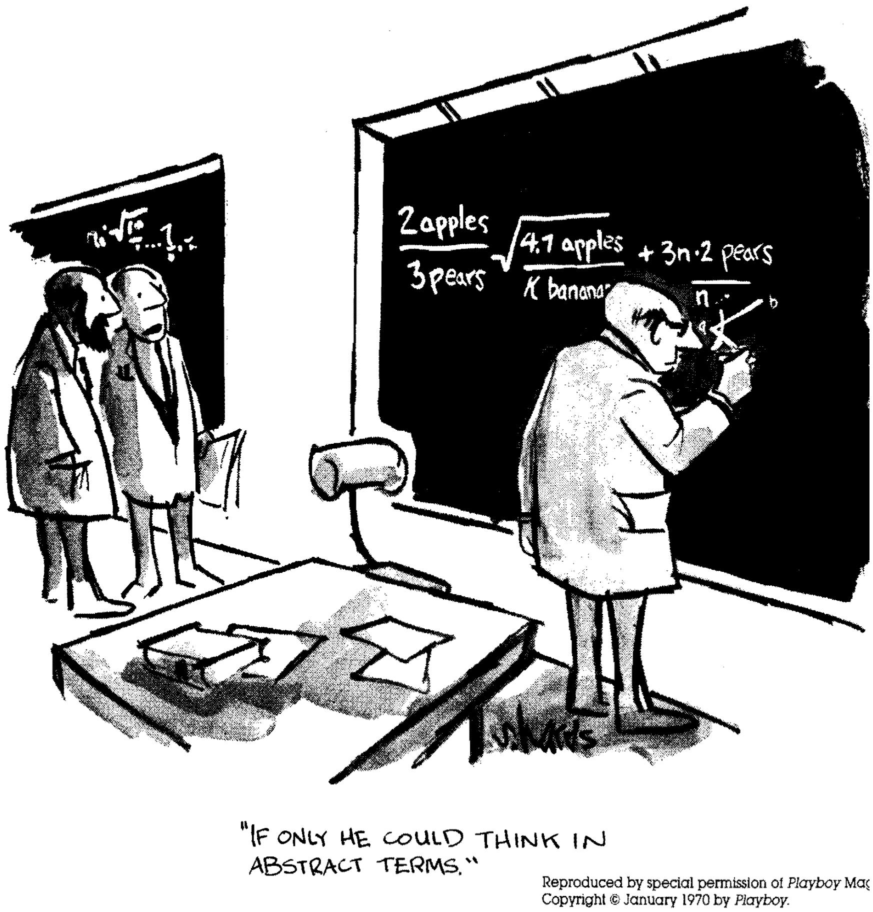

Figure 1.1: The power of abstraction

We can view the matrix A as a way of expressing a linear map from $\mathbb{R}^{3\times 1}$ to $\mathbb{R}^{3\times 1}$ and solving the system $Ax = b$ amounts to determining whether b belongs to the *image* (or *range*) of this linear map.

Yet another fruitful way of interpreting the resolution of the system $Ax = b$ is to view this problem as an *intersection problem*.

Indeed, each of the equations

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 1 \\ 2x_1 + x_2 + x_3 &= 2 \\ x_1 - 2x_2 - 2x_3 &= 3 \end{aligned}$$

defines a subset of $\mathbb{R}^3$ which is actually a *plane*.

The first equation

$$x_1 + 2x_2 - x_3 = 1$$

defines the plane H_1 passing through the three points $(1, 0, 0), (0, 1/2, 0), (0, 0, -1)$, on the coordinate axes, the second equation

$$2x_1 + x_2 + x_3 = 2$$

defines the plane H_2 passing through the three points $(1, 0, 0), (0, 2, 0), (0, 0, 2)$, on the coordinate axes, and the third equation

$$x_1 - 2x_2 - 2x_3 = 3$$

defines the plane H_3 passing through the three points $(3, 0, 0), (0, -3/2, 0), (0, 0, -3/2)$, on the coordinate axes.

The intersection $H_i \cap H_j$ of any two distinct planes H_i and H_j is a line, and the intersection $H_1 \cap H_2 \cap H_3$ of the three planes consists of the single point $(1.4, -0.4, -0.4)$.

Under this interpretation, observe that we are focusing on the *rows* of the matrix A, rather than on its *columns*, as in the previous interpretations.

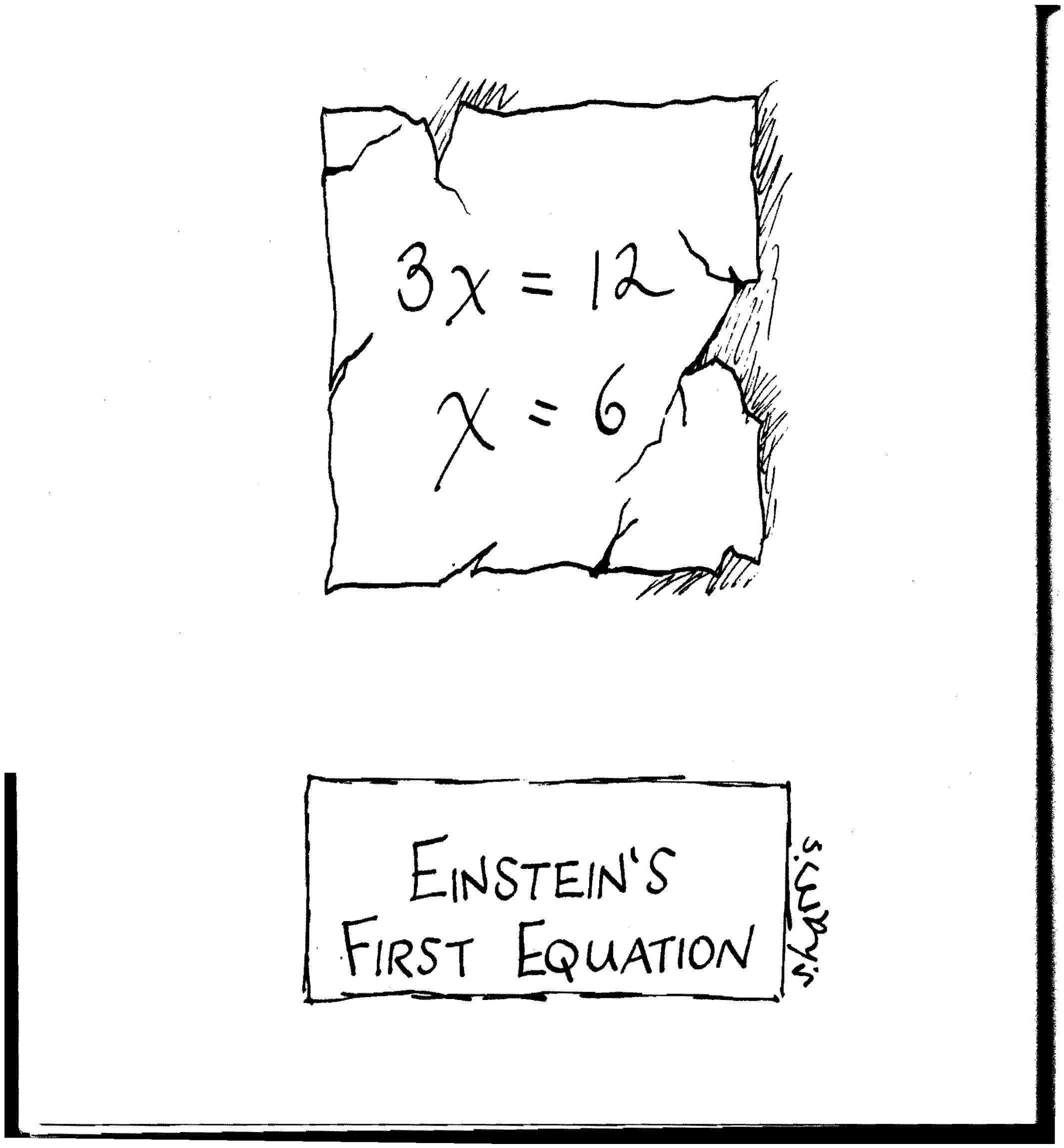

Figure 1.2: Linear Equations

Another great example of a real-world problem where linear algebra proves to be very effective is the problem of *data compression*, that is, of representing a very large data set using a much smaller amount of storage.

Typically the data set is represented as an $m \times n$ matrix A where each row corresponds to an n-dimensional data point and typically, $m \geq n$.

In most applications, the data are not independent so the rank of A is a lot smaller than $\min\{m, n\}$, and the the goal of *low-rank decomposition* is to factor A as the product of two matrices B and C, where B is a $m \times k$ matrix and C is a $k \times n$ matrix, with $k \ll \min\{m, n\}$ (here, $\ll$ means "much smaller than"):

$$\begin{pmatrix} & & \\ & A & \\ & m \times n & \\ & & \end{pmatrix} = \begin{pmatrix} & & \\ & B & \\ & m \times k & \\ & & \end{pmatrix} \begin{pmatrix} & C & \\ & k \times n & \end{pmatrix}$$

Now, it is generally too costly to find an exact factorization as above, so we look for a low-rank matrix A' which is a "good" *approximation* of A.

In order to make this statement precise, we need to define a mechanism to determine how close two matrices are. This can be done using *matrix norms*, a notion discussed in Chapter 6.

The norm of a matrix A is a nonnegative real number $\|A\|$ which behaves a lot like the absolute value $|x|$ of a real number x.

Then, our goal is to find some low-rank matrix A' that minimizes the norm

$$\|A - A'\|^2,$$

over all matrices A' of rank at most k, for some given $k \ll \min\{m, n\}$.

Some advantages of a low-rank approximation are:

1. Fewer elements are required to represent A; namely, $k(m+n)$ instead of mn. Thus less storage and fewer operations are needed to reconstruct A.
2. Often, the decomposition exposes the underlying structure of the data. Thus, it may turn out that "most" of the significant data are concentrated along some directions called *principal directions*.

Low-rank decompositions of a set of data have a multitude of applications in engineering, including computer science (especially computer vision), statistics, and machine learning.

As we will see later in Chapter 15, the *singular value decomposition* (SVD) provides a very satisfactory solution to the low-rank approximation problem.

Still, in many cases, the data sets are so large that another ingredient is needed: *randomization*. However, as a first step, linear algebra often yields a good initial solution.

We will now be more precise as to what kinds of operations are allowed on vectors.

In the early 1900, the notion of a *vector space* emerged as a convenient and unifying framework for working with "linear" objects.

1.2 Vector Spaces

A (real) vector space is a set E together with two operations, $+\colon E \times E \to E$ and $\cdot\colon \mathbb{R} \times E \to E$, called *addition* and *scalar multiplication*, that satisfy some simple properties.

First of all, E under addition has to be a commutative (or abelian) group, a notion that we review next.

However, keep in mind that vector spaces are not just algebraic objects; they are also geometric objects.

Definition 1.1. A *group* is a set G equipped with a binary operation $\cdot\colon G \times G \to G$ that associates an element $a \cdot b \in G$ to every pair of elements $a, b \in G$, and having the following properties: $\cdot$ is *associative*, has an *identity element* $e \in G$, and every element in G is *invertible* (w.r.t. $\cdot$).

More explicitly, this means that the following equations hold for all $a, b, c \in G$:

(G1) $a \cdot (b \cdot c) = (a \cdot b) \cdot c.$ (associativity);

(G2) $a \cdot e = e \cdot a = a.$ (identity);

(G3) For every $a \in G$, there is some $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$ (inverse).

A group G is *abelian* (or *commutative*) if

$$a \cdot b = b \cdot a$$

for all $a, b \in G$.

A set M together with an operation $\cdot\colon M \times M \to M$ and an element e satisfying only conditions (G1) and (G2) is called a *monoid*.

For example, the set $\mathbb{N} = \{0, 1, \ldots, n, \ldots\}$ of *natural numbers* is a (commutative) monoid under addition. However, it is not a group.

Example 1.1.

1. The set $\mathbb{Z} = \{\ldots, -n, \ldots, -1, 0, 1, \ldots, n, \ldots\}$ of *integers* is a group under addition, with identity element 0. However, $\mathbb{Z}^* = \mathbb{Z} - \{0\}$ is not a group under multiplication.

2. The set $\mathbb{Q}$ of *rational numbers* (fractions p/q with $p, q \in \mathbb{Z}$ and $q \neq 0$) is a group under addition, with identity element 0. The set $\mathbb{Q}^* = \mathbb{Q} - \{0\}$ is also a group under multiplication, with identity element 1.

3. Similarly, the sets $\mathbb{R}$ of *real numbers* and $\mathbb{C}$ of *complex numbers* are groups under addition (with identity element 0), and $\mathbb{R}^* = \mathbb{R} - \{0\}$ and $\mathbb{C}^* = \mathbb{C} - \{0\}$ are groups under multiplication (with identity element 1).

4. The sets $\mathbb{R}^n$ and $\mathbb{C}^n$ of n-tuples of real or complex numbers are groups under componentwise addition:

   $$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$

   with identity element $(0, \ldots, 0)$. All these groups are abelian.

5. Given any nonempty set S, the set of bijections $f \colon S \to S$, also called *permutations* of S, is a group under function composition (i.e., the multiplication of f and g is the composition $g \circ f$), with identity element the identity function id_S. This group is not abelian as soon as S has more than two elements.

6. The set of $n \times n$ matrices with real (or complex) coefficients is a group under addition of matrices, with identity element the null matrix. It is denoted by $\mathrm{M}_n(\mathbb{R})$ (or $\mathrm{M}_n(\mathbb{C})$).

7. The set $\mathbb{R}[X]$ of all polynomials in one variable with real coefficients is a group under addition of polynomials.

8. The set of $n \times n$ invertible matrices with real (or complex) coefficients is a group under matrix multiplication, with identity element the identity matrix I_n. This group is called the *general linear group* and is usually denoted by $\mathbf{GL}(n, \mathbb{R})$ (or $\mathbf{GL}(n, \mathbb{C})$).

9. The set of $n \times n$ invertible matrices with real (or complex) coefficients and determinant $+1$ is a group under matrix multiplication, with identity element the identity matrix I_n. This group is called the *special linear group* and is usually denoted by $\mathbf{SL}(n, \mathbb{R})$ (or $\mathbf{SL}(n, \mathbb{C})$).

10. The set of $n \times n$ invertible matrices with real coefficients such that $RR^\top = I_n$ and of determinant $+1$ is a group called the *special orthogonal group* and is usually denoted by $\mathbf{SO}(n)$ (where $R^\top$ is the *transpose* of the matrix R, i.e., the rows of $R^\top$ are the columns of R). It corresponds to the *rotations* in $\mathbb{R}^n$.

11. Given an open interval $]a, b[$, the set $\mathcal{C}(]a, b[)$ of continuous functions $f\colon]a, b[\to \mathbb{R}$ is a group under the operation $f + g$ defined such that

$$(f + g)(x) = f(x) + g(x)$$

for all $x \in]a, b[$.

It is customary to denote the operation of an abelian group G by $+$, in which case the inverse a^{-1} of an element $a \in G$ is denoted by $-a$.

Vector spaces are defined as follows.

Definition 1.2. A *real vector space* is a set E (of vectors) together with two operations $+\colon E \times E \to E$ (called *vector addition*)[1] and $\cdot\colon \mathbb{R} \times E \to E$ (called *scalar multiplication*) satisfying the following conditions for all $\alpha, \beta \in \mathbb{R}$ and all $u, v \in E$;

(V0) E is an abelian group w.r.t. $+$, with identity element 0;[2]

(V1) $\alpha \cdot (u + v) = (\alpha \cdot u) + (\alpha \cdot v)$;

(V2) $(\alpha + \beta) \cdot u = (\alpha \cdot u) + (\beta \cdot u)$;

(V3) $(\alpha * \beta) \cdot u = \alpha \cdot (\beta \cdot u)$;

(V4) $1 \cdot u = u$.

In (V3), $*$ denotes multiplication in $\mathbb{R}$.

Given $\alpha \in \mathbb{R}$ and $v \in E$, the element $\alpha \cdot v$ is also denoted by αv. The field $\mathbb{R}$ is often called the field of scalars.

In definition 1.2, the field $\mathbb{R}$ may be replaced by the field of complex numbers $\mathbb{C}$, in which case we have a *complex* vector space.

[1] The symbol $+$ is overloaded, since it denotes both addition in the field $\mathbb{R}$ and addition of vectors in E. It is usually clear from the context which $+$ is intended.

[2] The symbol 0 is also overloaded, since it represents both the zero in $\mathbb{R}$ (a scalar) and the identity element of E (the zero vector). Confusion rarely arises, but one may prefer using $\mathbf{0}$ for the zero vector.

It is even possible to replace $\mathbb{R}$ by the field of rational numbers $\mathbb{Q}$ or by any other field K (for example $\mathbb{Z}/p\mathbb{Z}$, where p is a prime number), in which case we have a *K-vector space* (in (V3), $*$ denotes multiplication in the field K).

In most cases, the field K will be the field $\mathbb{R}$ of reals.

From (V0), a vector space always contains the null vector 0, and thus is nonempty.

From (V1), we get $\alpha \cdot 0 = 0$, and $\alpha \cdot (-v) = -(\alpha \cdot v)$.

From (V2), we get $0 \cdot v = 0$, and $(-\alpha) \cdot v = -(\alpha \cdot v)$.

Another important consequence of the axioms is the following fact: For any $u \in E$ and any $\lambda \in \mathbb{R}$, if $\lambda \neq 0$ and $\lambda \cdot u = 0$, then $u = 0$.

The field $\mathbb{R}$ itself can be viewed as a vector space over itself, addition of vectors being addition in the field, and multiplication by a scalar being multiplication in the field.

Example 1.2.

1. The fields $\mathbb{R}$ and $\mathbb{C}$ are vector spaces over $\mathbb{R}$.
2. The groups $\mathbb{R}^n$ and $\mathbb{C}^n$ are vector spaces over $\mathbb{R}$, and $\mathbb{C}^n$ is a vector space over $\mathbb{C}$.
3. The ring $\mathbb{R}[X]_n$ of polynomials of degree at most n with real coefficients is a vector space over $\mathbb{R}$, and the ring $\mathbb{C}[X]_n$ of polynomials of degree at most n with complex coefficients is a vector space over $\mathbb{C}$.
4. The ring $\mathbb{R}[X]$ of all polynomials with real coefficients is a vector space over $\mathbb{R}$, and the ring $\mathbb{C}[X]$ of all polynomials with complex coefficients is a vector space over $\mathbb{C}$.
5. The ring of $n \times n$ matrices $\mathrm{M}_n(\mathbb{R})$ is a vector space over $\mathbb{R}$.
6. The ring of $m \times n$ matrices $\mathrm{M}_{m,n}(\mathbb{R})$ is a vector space over $\mathbb{R}$.
7. The ring $\mathcal{C}(]a, b[)$ of continuous functions $f\colon]a, b[\to \mathbb{R}$ is a vector space over $\mathbb{R}$.

Let E be a vector space. We would like to define the important notions of linear combination and linear independence.

These notions can be defined for sets of vectors in E, but it will turn out to be more convenient to define them for families $(v_i)_{i \in I}$, where I is any arbitrary index set.

1.3 Linear Independence, Subspaces

One of the most useful properties of vector spaces is that there possess bases.

What this means is that in every vector space, E, there is some set of vectors, $\{e_1, \ldots, e_n\}$, such that *every* vector $v \in E$ can be written as a linear combination,

$$v = \lambda_1 e_1 + \cdots + \lambda_n e_n,$$

of the e_i, for some scalars, $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.

Furthermore, the n-tuple, $(\lambda_1, \ldots, \lambda_n)$, as above is *unique*.

This description is fine when E has a finite basis, $\{e_1, \ldots, e_n\}$, but this is not always the case!

For example, the vector space of real polynomials, $\mathbb{R}[X]$, does not have a finite basis but instead it has an infinite basis, namely

$$1,\ X,\ X^2,\ \ldots, X^n,\ \ldots$$

For simplicity, in this chapter, we will restrict our attention to vector spaces that have a finite basis (we say that they are *finite-dimensional*).

Given a set A, an *I-indexed family* $(a_i)_{i\in I}$ of elements of A (for short, a *family*) is simply a function $a\colon I \to A$, or equivalently a set of pairs $\{(i, a_i) \mid i \in I\}$.

We agree that when $I = \emptyset$, $(a_i)_{i\in I} = \emptyset$. A family $(a_i)_{i\in I}$ is finite if I is finite.

Remark: When considering a family $(a_i)_{i\in I}$, there is no reason to assume that I is ordered.

The crucial point is that every element of the family is uniquely indexed by an element of I.

Thus, unless specified otherwise, we do not assume that the elements of an index set are ordered.

Given a family $(u_i)_{i \in I}$ and any element v, we denote by

$$(u_i)_{i \in I} \cup_k (v)$$

the family $(w_i)_{i \in I \cup \{k\}}$ defined such that, $w_i = u_i$ if $i \in I$, and $w_k = v$, where k is any index such that $k \notin I$.

Given a family $(u_i)_{i \in I}$, a *subfamily* of $(u_i)_{i \in I}$ is a family $(u_j)_{j \in J}$ where J is any subset of I.

In this chapter, unless specified otherwise, it is assumed that all families of scalars are *finite* (i.e., their index set is finite).

Definition 1.3. Let E be a vector space. A vector $v \in E$ is a *linear combination of a family $(u_i)_{i\in I}$ of elements of E* iff there is a family $(\lambda_i)_{i\in I}$ of scalars in $\mathbb{R}$ such that

$$v = \sum_{i\in I} \lambda_i u_i.$$

When $I = \emptyset$, we stipulate that $v = 0$.

We say that a family $(u_i)_{i\in I}$ is *linearly independent* iff for every family $(\lambda_i)_{i\in I}$ of scalars in $\mathbb{R}$,

$$\sum_{i\in I} \lambda_i u_i = 0 \quad \text{implies that} \quad \lambda_i = 0 \ \text{ for all } i \in I.$$

Equivalently, a family $(u_i)_{i\in I}$ is *linearly dependent* iff there is some family $(\lambda_i)_{i\in I}$ of scalars in $\mathbb{R}$ such that

$$\sum_{i\in I} \lambda_i u_i = 0 \quad \text{and} \quad \lambda_j \neq 0 \ \text{ for some } j \in I.$$

We agree that when $I = \emptyset$, the family $\emptyset$ is linearly independent.

A family $(u_i)_{i\in I}$ is linearly independent iff either $I = \emptyset$, or I consists of a single element i and $u_i \neq 0$, or $|I| \geq 2$ and no vector u_j in the family can be expressed as a linear combination of the other vectors in the family.

A family $(u_i)_{i \in I}$ is linearly dependent iff either I consists of a single element, say i, and $u_i = 0$, or $|I| \geq 2$ and some u_j in the family can be expressed as a linear combination of the other vectors in the family.

When I is nonempty, if the family $(u_i)_{i \in I}$ is linearly independent, then $u_i \neq 0$ for all $i \in I$. Furthermore, if $|I| \geq 2$, then $u_i \neq u_j$ for all $i, j \in I$ with $i \neq j$.

Example 1.3.

1. Any two distinct scalars $\lambda, \mu \neq 0$ in $\mathbb{R}$ are linearly dependent.
2. In $\mathbb{R}^3$, the vectors $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ are linearly independent.
3. In $\mathbb{R}^4$, the vectors $(1, 1, 1, 1)$, $(0, 1, 1, 1)$, $(0, 0, 1, 1)$, and $(0, 0, 0, 1)$ are linearly independent.
4. In $\mathbb{R}^2$, the vectors $u = (1, 1)$, $v = (0, 1)$ and $w = (2, 3)$ are linearly dependent, since
$$w = 2u + v.$$

When I is finite, we often assume that it is the set $I = \{1, 2, \ldots, n\}$. In this case, we denote the family $(u_i)_{i \in I}$ as $(u_1, \ldots, u_n)$.

The notion of a subspace of a vector space is defined as follows.

Definition 1.4. Given a vector space E, a subset F of E is a *linear subspace* (or *subspace*) of E iff F is nonempty and $\lambda u + \mu v \in F$ for all $u, v \in F$, and all $\lambda, \mu \in \mathbb{R}$.

It is easy to see that a subspace F of E is indeed a vector space.

It is also easy to see that any *intersection* of subspaces is a subspace.

Every subspace contains the vector 0.

For any nonempty finite index set I, one can show by induction on the cardinality of I that if $(u_i)_{i \in I}$ is any family of vectors $u_i \in F$ and $(\lambda_i)_{i \in I}$ is any family of scalars, then $\sum_{i \in I} \lambda_i u_i \in F$.

The subspace $\{0\}$ will be denoted by (0), or even 0 (with a mild abuse of notation).

Example 1.4.

1. In $\mathbb{R}^2$, the set of vectors $u = (x, y)$ such that
$$x + y = 0$$
is a subspace.
2. In $\mathbb{R}^3$, the set of vectors $u = (x, y, z)$ such that
$$x + y + z = 0$$
is a subspace.
3. For any $n \geq 0$, the set of polynomials $f(X) \in \mathbb{R}[X]$ of degree at most n is a subspace of $\mathbb{R}[X]$.
4. The set of upper triangular $n \times n$ matrices is a subspace of the space of $n \times n$ matrices.

Proposition 1.1. *Given any vector space E, if S is any nonempty subset of E, then the smallest subspace $\langle S \rangle$ (or $\mathrm{Span}(S)$) of E containing S is the set of all (finite) linear combinations of elements from S.*

One might wonder what happens if we add extra conditions to the coefficients involved in forming linear combinations.

Here are three natural restrictions which turn out to be important (as usual, we assume that our index sets are finite):

(1) Consider combinations $\sum_{i\in I} \lambda_i u_i$ for which

$$\sum_{i\in I} \lambda_i = 1.$$

These are called *affine combinations*.

One should realize that every linear combination $\sum_{i\in I} \lambda_i u_i$ can be viewed as an affine combination.

However, we get new spaces. For example, in $\mathbb{R}^3$, the set of all affine combinations of the three vectors $e_1 = (1, 0, 0), e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$, is the plane passing through these three points.

Since it does not contain $0 = (0, 0, 0)$, it is not a linear subspace.

(2) Consider combinations $\sum_{i \in I} \lambda_i u_i$ for which

$$\lambda_i \geq 0, \quad \text{for all } i \in I.$$

These are called *positive* (or *conic*) *combinations*.

It turns out that positive combinations of families of vectors are *cones*. They show up naturally in convex optimization.

(3) Consider combinations $\sum_{i \in I} \lambda_i u_i$ for which we require (1) *and* (2), that is

$$\sum_{i \in I} \lambda_i = 1, \quad \text{and} \quad \lambda_i \geq 0 \quad \text{for all } i \in I.$$

These are called *convex combinations*.

Given any finite family of vectors, the set of all convex combinations of these vectors is a *convex polyhedron*.

Convex polyhedra play a very important role in *convex optimization*.

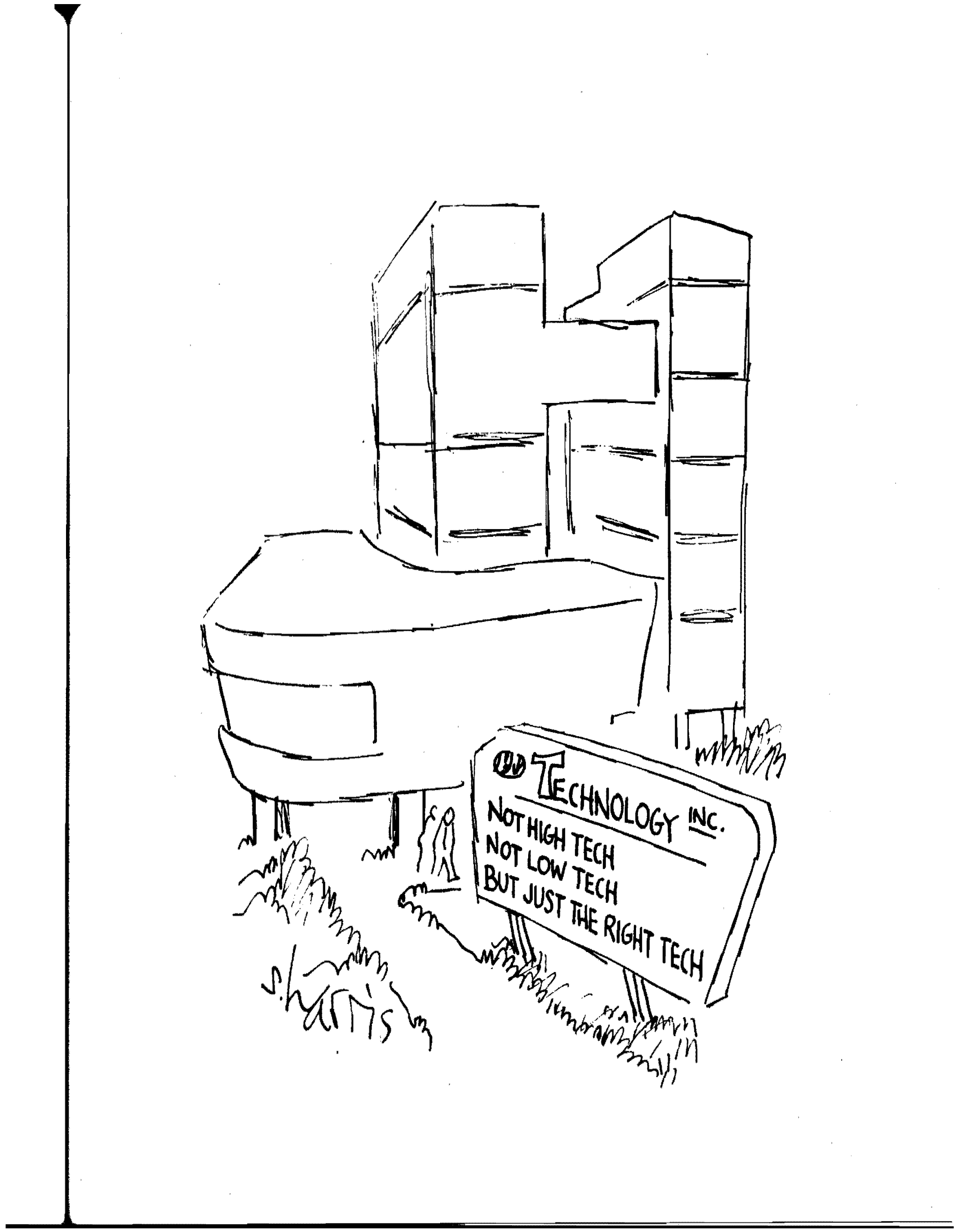

Figure 1.3: The right Tech

1.4 Bases of a Vector Space

Definition 1.5. Given a vector space E and a subspace V of E, a family $(v_i)_{i \in I}$ of vectors $v_i \in V$ *spans V* or *generates V* iff for every $v \in V$, there is some family $(\lambda_i)_{i \in I}$ of scalars in $\mathbb{R}$ such that

$$v = \sum_{i \in I} \lambda_i v_i.$$

We also say that the elements of $(v_i)_{i \in I}$ are *generators* of V and that V is *spanned by $(v_i)_{i \in I}$*, or *generated by $(v_i)_{i \in I}$*.

If a subspace V of E is generated by a finite family $(v_i)_{i \in I}$, we say that V is *finitely generated*.

A family $(u_i)_{i \in I}$ that spans V and is linearly independent is called a *basis* of V.

Example 1.5.

1. In $\mathbb{R}^3$, the vectors $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ form a basis.

2. The vectors $(1,1,1,1), (1,1,-1,-1), (1,-1,0,0)$, $(0,0,1,-1)$ form a basis of $\mathbb{R}^4$ known as the *Haar basis*. This basis and its generalization to dimension 2^n are crucial in wavelet theory.

3. In the subspace of polynomials in $\mathbb{R}[X]$ of degree at most n, the polynomials $1, X, X^2, \ldots, X^n$ form a basis.

4. The *Bernstein polynomials* $\binom{n}{k}(1-X)^{n-k}X^k$ for $k = 0, \ldots, n$, also form a basis of that space. These polynomials play a major role in the theory of *spline curves*.

It is a standard result of linear algebra that every vector space E has a basis, and that for any two bases $(u_i)_{i\in I}$ and $(v_j)_{j\in J}$, I and J have the same cardinality.

In particular, if E has a finite basis of n elements, every basis of E has n elements, and the integer n is called the *dimension* of the vector space E.

We begin with a crucial lemma.

Lemma 1.2. *Given a linearly independent family* $(u_i)_{i\in I}$ *of elements of a vector space* E*, if* $v \in E$ *is not a linear combination of* $(u_i)_{i\in I}$*, then the family* $(u_i)_{i\in I} \cup_k (v)$ *obtained by adding* v *to the family* $(u_i)_{i\in I}$ *is linearly independent (where* $k \notin I$*).*

The next theorem holds in general, but the proof is more sophisticated for vector spaces that do not have a finite set of generators.

Theorem 1.3. *Given any finite family* $S = (u_i)_{i\in I}$ *generating a vector space* E *and any linearly independent subfamily* $L = (u_j)_{j\in J}$ *of* S *(where* $J \subseteq I$*), there is a basis* B *of* E *such that* $L \subseteq B \subseteq S$*.*

Let $(v_i)_{i\in I}$ be a family of vectors in E. We say that $(v_i)_{i\in I}$ a *maximal linearly independent family of* E if it is linearly independent, and if for any vector $w \in E$, the family $(v_i)_{i\in I} \cup_k \{w\}$ obtained by adding w to the family $(v_i)_{i\in I}$ is linearly dependent.

We say that $(v_i)_{i\in I}$ a *minimal generating family of E* if it spans E, and if for any index $p \in I$, the family $(v_i)_{i\in I-\{p\}}$ obtained by removing v_p from the family $(v_i)_{i\in I}$ does not span E.

The following proposition giving useful properties characterizing a basis is an immediate consequence of Theorem 1.3.

Proposition 1.4. *Given a vector space E, for any family $B = (v_i)_{i\in I}$ of vectors of E, the following properties are equivalent:*

(1) B is a basis of E.

(2) B is a maximal linearly independent family of E.

(3) B is a minimal generating family of E.

The following *replacement lemma* due to Steinitz shows the relationship between finite linearly independent families and finite families of generators of a vector space.

We begin with a version of the lemma which is a bit informal, but easier to understand than the precise and more formal formulation given in Proposition 1.6. The technical difficulty has to do with the fact that some of the indices need to be renamed.

Proposition 1.5. *(Replacement lemma, version 1) Given a vector space E, let $(u_1, \ldots, u_m)$ be any finite linearly independent family in E, and let $(v_1, \ldots, v_n)$ be any finite family such that every u_i is a linear combination of $(v_1, \ldots, v_n)$. Then, we must have $m \leq n$, and m of the vectors v_j can be replaced by $(u_1, \ldots, u_m)$, such that after renaming some of the indices of the vs, the families $(u_1, \ldots, u_m, v_{m+1}, \ldots, v_n)$ and $(v_1, \ldots, v_n)$ generate the same subspace of E.*

The idea is that m of the vectors v_j can be *replaced* by the linearly independent u_i's in such a way that the same subspace is still generated.

Proposition 1.6. *(Replacement lemma, version 2) Given a vector space E, let $(u_i)_{i \in I}$ be any finite linearly independent family in E, where $|I| = m$, and let $(v_j)_{j \in J}$ be any finite family such that every u_i is a linear combination of $(v_j)_{j \in J}$, where $|J| = n$. Then, there exists a set L and an injection $\rho\colon L \to J$ (a relabeling function) such that $L \cap I = \emptyset$, $|L| = n - m$, and the families $(u_i)_{i \in I} \cup (v_{\rho(l)})_{l \in L}$ and $(v_j)_{j \in J}$ generate the same subspace of E. In particular, $m \leq n$.*

The purpose of the function $\rho\colon L \to J$ is to pick $n-m$ elements $j_1, \ldots, j_{n-m}$ of J and to relabel them $l_1, \ldots, l_{n-m}$ in such a way that these new indices do not clash with the indices in I; this way, the vectors $v_{j_1}, \ldots, v_{j_{n-m}}$ who "survive" (i.e. are not replaced) are relabeled $v_{l_1}, \ldots, v_{l_{n-m}}$, and the other m vectors v_j with $j \in J - \{j_1, \ldots, j_{n-m}\}$ are replaced by the u_i. The index set of this new family is $I \cup L$.

Actually, one can prove that Proposition 1.6 implies Theorem 1.3 when the vector space is finitely generated.

Putting Theorem 1.3 and Proposition 1.6 together, we obtain the following fundamental theorem.

Theorem 1.7. *Let E be a finitely generated vector space. Any family $(u_i)_{i \in I}$ generating E contains a subfamily $(u_j)_{j \in J}$ which is a basis of E. Any linearly independent family $(u_i)_{i \in I}$ can be extended to a family $(u_j)_{j \in J}$ which is a basis of E (with $I \subseteq J$). Furthermore, for every two bases $(u_i)_{i \in I}$ and $(v_j)_{j \in J}$ of E, we have $|I| = |J| = n$ for some fixed integer $n \geq 0$.*

Remark: Theorem 1.7 also holds for vector spaces that are not finitely generated.

When E is not finitely generated, we say that E is of *infinite dimension*.

The *dimension* of a finitely generated vector space E is the common dimension n of all of its bases and is denoted by $\dim(E)$.

Clearly, if the field $\mathbb{R}$ itself is viewed as a vector space, then every family (a) where $a \in \mathbb{R}$ and $a \neq 0$ is a basis. Thus $\dim(\mathbb{R}) = 1$.

Note that $\dim(\{0\}) = 0$.

If E is a vector space of dimension $n \geq 1$, for any subspace U of E,

if $\dim(U) = 1$, then U is called a *line*;

if $\dim(U) = 2$, then U is called a *plane*;

if $\dim(U) = n - 1$, then U is called a *hyperplane*.

If $\dim(U) = k$, then U is sometimes called a *k-plane*.

Let $(u_i)_{i \in I}$ be a *basis* of a vector space E.

For any vector $v \in E$, since the family $(u_i)_{i \in I}$ generates E, there is a family $(\lambda_i)_{i \in I}$ of scalars in $\mathbb{R}$, such that

$$v = \sum_{i \in I} \lambda_i u_i.$$

A very important fact is that the family $(\lambda_i)_{i\in I}$ is *unique*.

Proposition 1.8. *Given a vector space E, let $(u_i)_{i\in I}$ be a family of vectors in E. Let $v \in E$, and assume that $v = \sum_{i\in I} \lambda_i u_i$. Then, the family $(\lambda_i)_{i\in I}$ of scalars such that $v = \sum_{i\in I} \lambda_i u_i$ is unique iff $(u_i)_{i\in I}$ is linearly independent.*

If $(u_i)_{i\in I}$ is a basis of a vector space E, for any vector $v \in E$, if $(x_i)_{i\in I}$ is the unique family of scalars in $\mathbb{R}$ such that

$$v = \sum_{i\in I} x_i u_i,$$

each x_i is called the *component (or coordinate) of index i of v with respect to the basis $(u_i)_{i\in I}$*.

Many interesting mathematical structures are vector spaces.

A very important example is the set of linear maps between two vector spaces to be defined in the next section.

Here is an example that will prepare us for the vector space of linear maps.

Example 1.6. Let X be any nonempty set and let E be a vector space. The set of all functions $f\colon X \to E$ can be made into a vector space as follows: Given any two functions $f\colon X \to E$ and $g\colon X \to E$, let $(f+g)\colon X \to E$ be defined such that

$$(f+g)(x) = f(x) + g(x)$$

for all $x \in X$, and for every $\lambda \in \mathbb{R}$, let $\lambda f\colon X \to E$ be defined such that

$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in X$.

The axioms of a vector space are easily verified.

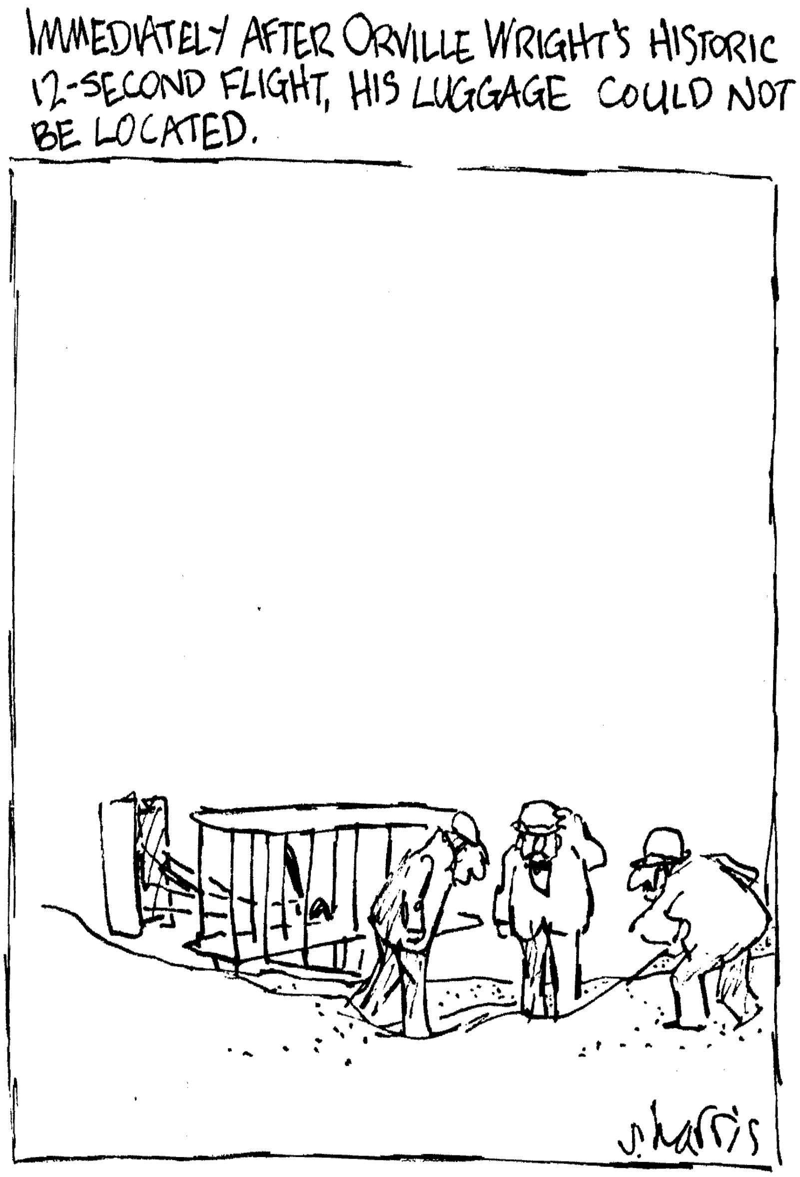

Figure 1.4: Early Traveling

1.5 Linear Maps

A function between two vector spaces that preserves the vector space structure is called a homomorphism of vector spaces, or linear map.

Linear maps formalize the concept of linearity of a function.

Keep in mind that linear maps, which are transformations of space, are usually far more important than the spaces themselves.

In the rest of this section, we assume that all vector spaces are real vector spaces.

Definition 1.6. Given two vector spaces E and F, a *linear map* between E and F is a function $f\colon E \to F$ satisfying the following two conditions:

$$\begin{aligned} f(x+y) &= f(x)+f(y) && \text{for all } x, y \in E; \\ f(\lambda x) &= \lambda f(x) && \text{for all } \lambda \in \mathbb{R},\ x \in E. \end{aligned}$$

Setting $x = y = 0$ in the first identity, we get $f(0) = 0$.

The basic property of linear maps is that they transform linear combinations into linear combinations.

Given any finite family $(u_i)_{i\in I}$ of vectors in E, given any family $(\lambda_i)_{i\in I}$ of scalars in $\mathbb{R}$, we have

$$f(\sum_{i\in I}\lambda_i u_i) = \sum_{i\in I}\lambda_i f(u_i).$$

The above identity is shown by induction on $|I|$ using the properties of Definition 1.6.

Example 1.7.

1. The map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined such that
$$\begin{aligned} x' &= x - y \\ y' &= x + y \end{aligned}$$
is a linear map.

2. For any vector space E, the *identity map* $\mathrm{id}\colon E \to E$ given by
$$\mathrm{id}(u) = u \quad \text{for all } u \in E$$
is a linear map. When we want to be more precise, we write id_E instead of id.

3. The map $D\colon \mathbb{R}[X] \to \mathbb{R}[X]$ defined such that
$$D(f(X)) = f'(X),$$
where $f'(X)$ is the derivative of the polynomial $f(X)$, is a linear map

Definition 1.7. Given a linear map $f\colon E \to F$, we define its *image (or range)* $\operatorname{Im} f = f(E)$, as the set

$$\operatorname{Im} f = \{y \in F \mid (\exists x \in E)(y = f(x))\},$$

and its *Kernel (or nullspace)* $\operatorname{Ker} f = f^{-1}(0)$, as the set

$$\operatorname{Ker} f = \{x \in E \mid f(x) = 0\}.$$

Proposition 1.9. *Given a linear map $f\colon E \to F$, the set $\operatorname{Im} f$ is a subspace of F and the set $\operatorname{Ker} f$ is a subspace of E. The linear map $f\colon E \to F$ is injective iff $\operatorname{Ker} f = 0$ (where 0 is the trivial subspace $\{0\}$).*

Since by Proposition 1.9, the image $\operatorname{Im} f$ of a linear map f is a subspace of F, we can define the *rank* $\operatorname{rk}(f)$ of f as the dimension of $\operatorname{Im} f$.

A fundamental property of bases in a vector space is that they allow the definition of linear maps as unique homomorphic extensions, as shown in the following proposition.

Proposition 1.10. *Given any two vector spaces E and F, given any basis $(u_i)_{i\in I}$ of E, given any other family of vectors $(v_i)_{i\in I}$ in F, there is a unique linear map $f\colon E \to F$ such that $f(u_i) = v_i$ for all $i \in I$.*

Furthermore, f is injective iff $(v_i)_{i\in I}$ is linearly independent, and f is surjective iff $(v_i)_{i\in I}$ generates F.

By the second part of Proposition 1.10, an injective linear map $f\colon E \to F$ sends a basis $(u_i)_{i\in I}$ to a linearly independent family $(f(u_i))_{i\in I}$ of F, which is also a basis when f is bijective.

Also, when E and F have the same finite dimension n, $(u_i)_{i\in I}$ is a basis of E, and $f\colon E \to F$ is injective, then $(f(u_i))_{i\in I}$ is a basis of F (by Proposition 1.4).

The following simple proposition is also useful.

Proposition 1.11. *Given any two vector spaces E and F, with F nontrivial, given any family $(u_i)_{i\in I}$ of vectors in E, the following properties hold:*

(1) The family $(u_i)_{i\in I}$ generates E iff for every family of vectors $(v_i)_{i\in I}$ in F, there is at most one linear map $f\colon E \to F$ such that $f(u_i) = v_i$ for all $i \in I$.

(2) The family $(u_i)_{i\in I}$ is linearly independent iff for every family of vectors $(v_i)_{i\in I}$ in F, there is some linear map $f\colon E \to F$ such that $f(u_i) = v_i$ for all $i \in I$.

Given vector spaces E, F, and G, and linear maps $f\colon E \to F$ and $g\colon F \to G$, it is easily verified that the composition $g \circ f\colon E \to G$ of f and g is a linear map.

A linear map $f\colon E \to F$ is an *isomorphism* iff there is a linear map $g\colon F \to E$, such that

$$g \circ f = \mathrm{id}_E \quad \text{and} \quad f \circ g = \mathrm{id}_F. \tag{$*$}$$

It is immediately verified that such a map g is unique.

The map g is called the *inverse* of f and it is also denoted by f^{-1}.

Proposition 1.10 shows that if $F = \mathbb{R}^n$, then we get an isomorphism between any vector space E of dimension $|J| = n$ and $\mathbb{R}^n$.

One can verify that if $f\colon E \to F$ is a bijective linear map, then its inverse $f^{-1}\colon F \to E$ is also a linear map, and thus f is an isomorphism.

Another useful corollary of Proposition 1.10 is this:

Proposition 1.12. *Let E be a vector space of finite dimension $n \geq 1$ and let $f \colon E \to E$ be any linear map. The following properties hold:*

(1) If f has a left inverse g, that is, if g is a linear map such that $g \circ f = \mathrm{id}$, then f is an isomorphism and $f^{-1} = g$.

(2) If f has a right inverse h, that is, if h is a linear map such that $f \circ h = \mathrm{id}$, then f is an isomorphism and $f^{-1} = h$.

The *set of all linear maps between two vector spaces E and F* is denoted by $\mathrm{Hom}(E, F)$.

When we wish to be more precise and specify the field K over which the vector spaces E and F are defined we write $\mathrm{Hom}_K(E, F)$.

The set $\mathrm{Hom}(E, F)$ is a vector space under the operations defined at the end of Section 1.1, namely

$$(f + g)(x) = f(x) + g(x)$$

for all $x \in E$, and

$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in E$.

When E and F have finite dimensions, the vector space $\mathrm{Hom}(E, F)$ also has finite dimension, as we shall see shortly.

When $E = F$, a linear map $f \colon E \to E$ is also called an *endomorphism*. The space $\mathrm{Hom}(E, E)$ is also denoted by $\mathrm{End}(E)$.

It is also important to note that composition confers to $\mathrm{Hom}(E, E)$ a ring structure.

Indeed, composition is an operation $\circ\colon \mathrm{Hom}(E, E) \times \mathrm{Hom}(E, E) \to \mathrm{Hom}(E, E)$, which is associative and has an identity id_E, and the distributivity properties hold:

$$\begin{aligned}(g_1 + g_2) \circ f &= g_1 \circ f + g_2 \circ f;\\ g \circ (f_1 + f_2) &= g \circ f_1 + g \circ f_2.\end{aligned}$$

The ring $\mathrm{Hom}(E, E)$ is an example of a noncommutative ring.

It is easily seen that the set of bijective linear maps $f\colon E \to E$ is a *group* under composition. Bijective linear maps are also called *automorphisms*.

The group of automorphisms of E is called the *general linear group (of E)*, and it is denoted by $\mathbf{GL}(E)$, or by $\mathrm{Aut}(E)$, or when $E = \mathbb{R}^n$, by $\mathbf{GL}(n, \mathbb{R})$, or even by $\mathbf{GL}(n)$.

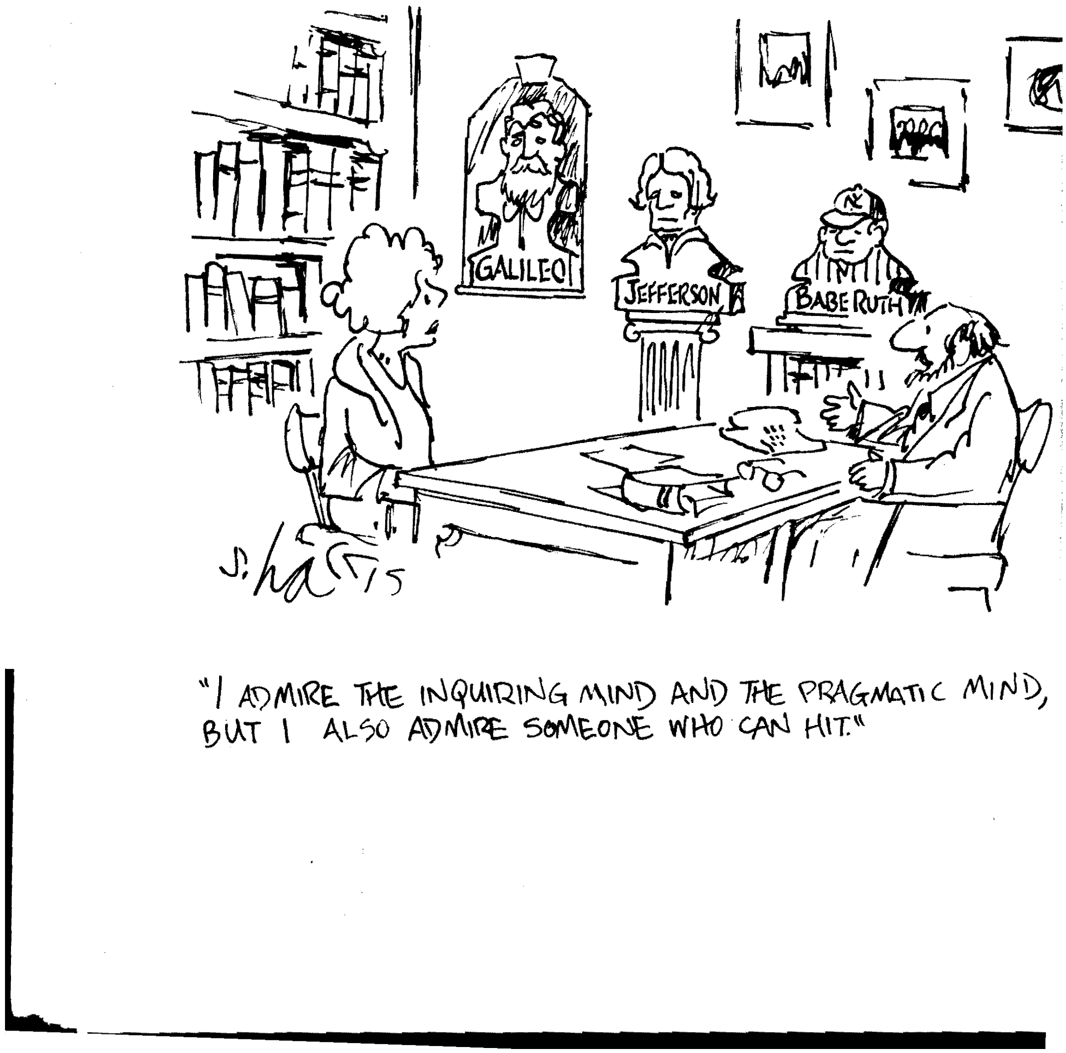

Figure 1.5: Hitting Power