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

November 5, 2013

Contents

Chapter 1

Basics of Linear Algebra

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

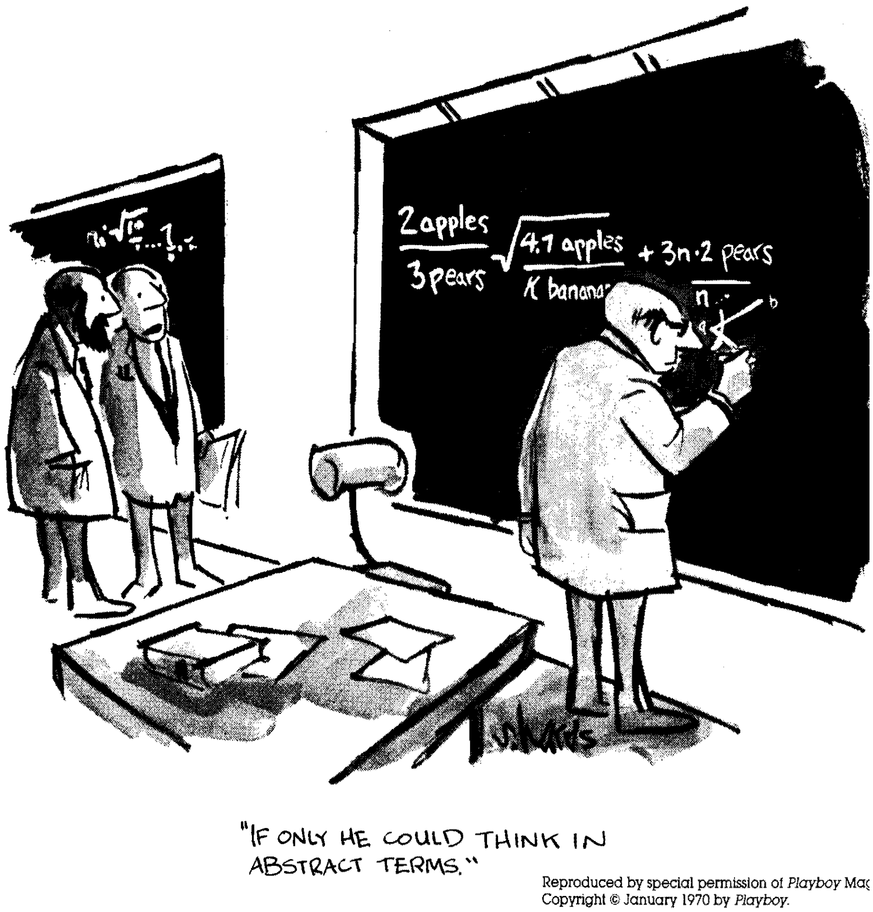


Figure 1.1: The power of abstraction

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.

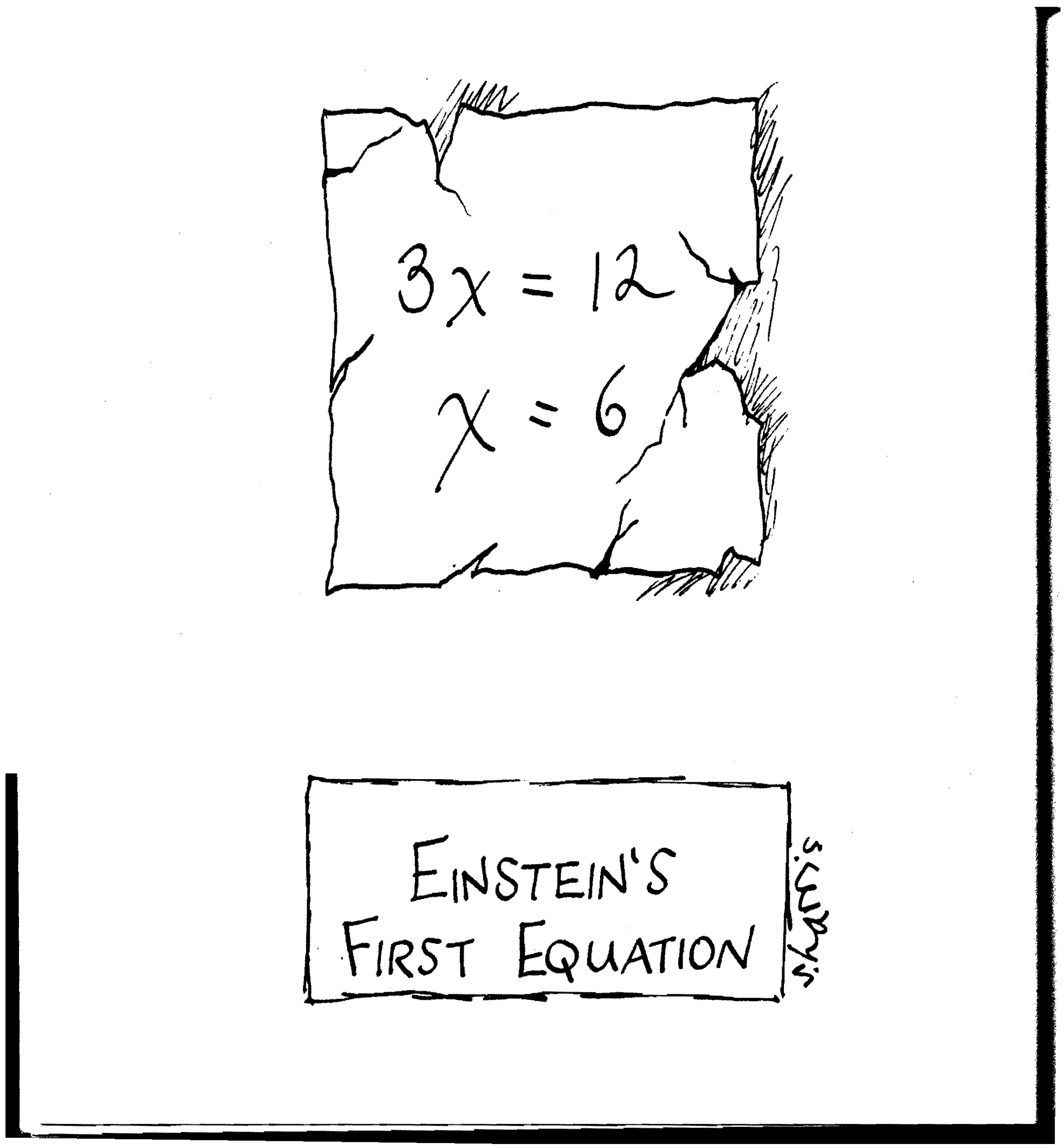


Figure 1.2: Linear Equations

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 4.

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 13, 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 $a\colon I \to A$.

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.

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.

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*.

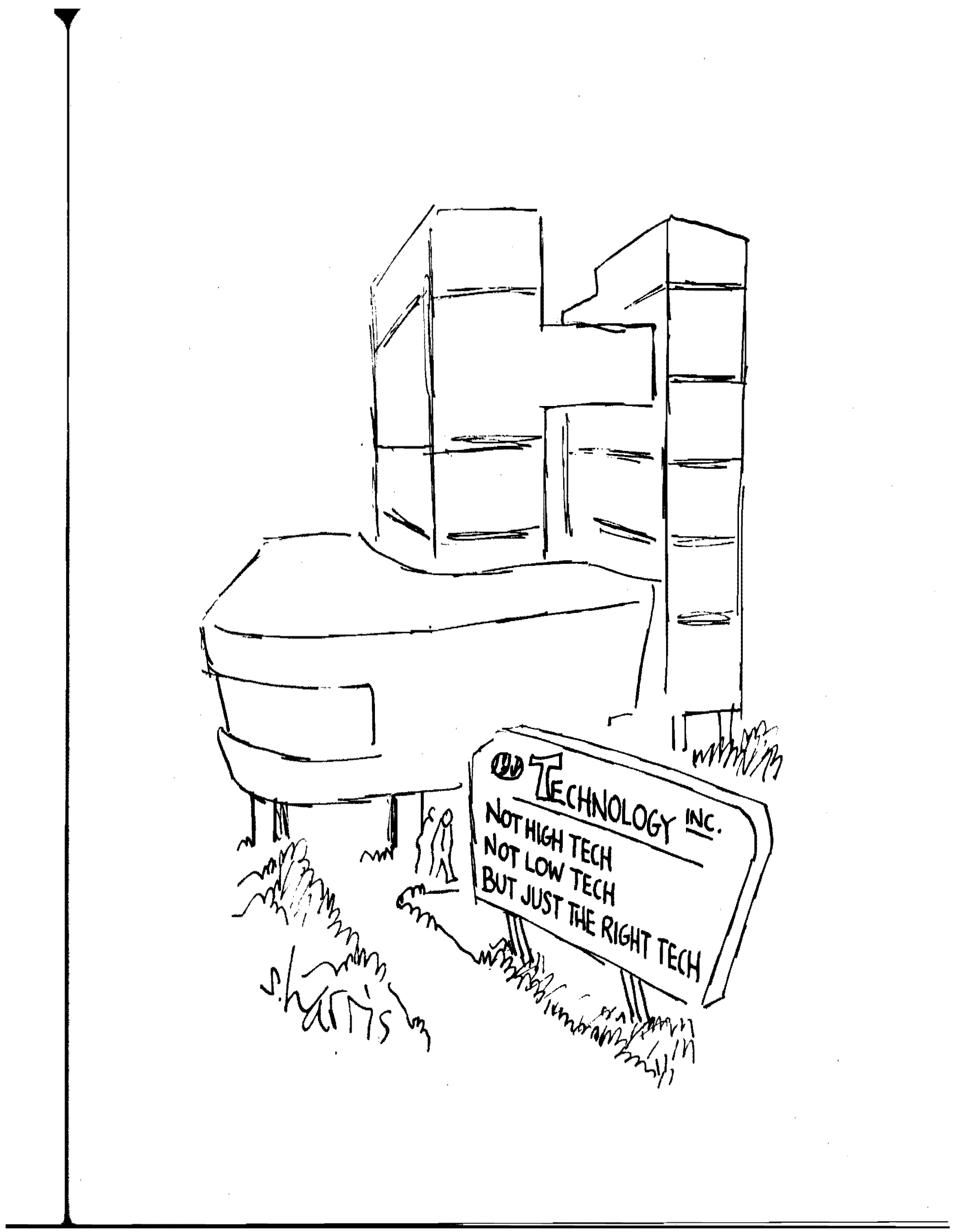


Figure 1.3: The right Tech

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* $\binom{n}{k}(1-X)^k X^{n-k}$ 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* $L \subseteq B \subseteq S$*.*

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.

Proposition 1.5. *(Replacement lemma) 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 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.

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 1.5 implies Theorem 1.3 when the vector space is finitely generated.

Putting Theorem 1.3 and Proposition 1.5 together, we obtain the following fundamental theorem.

Theorem 1.6. *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. 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 1.6 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 1.7. *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.

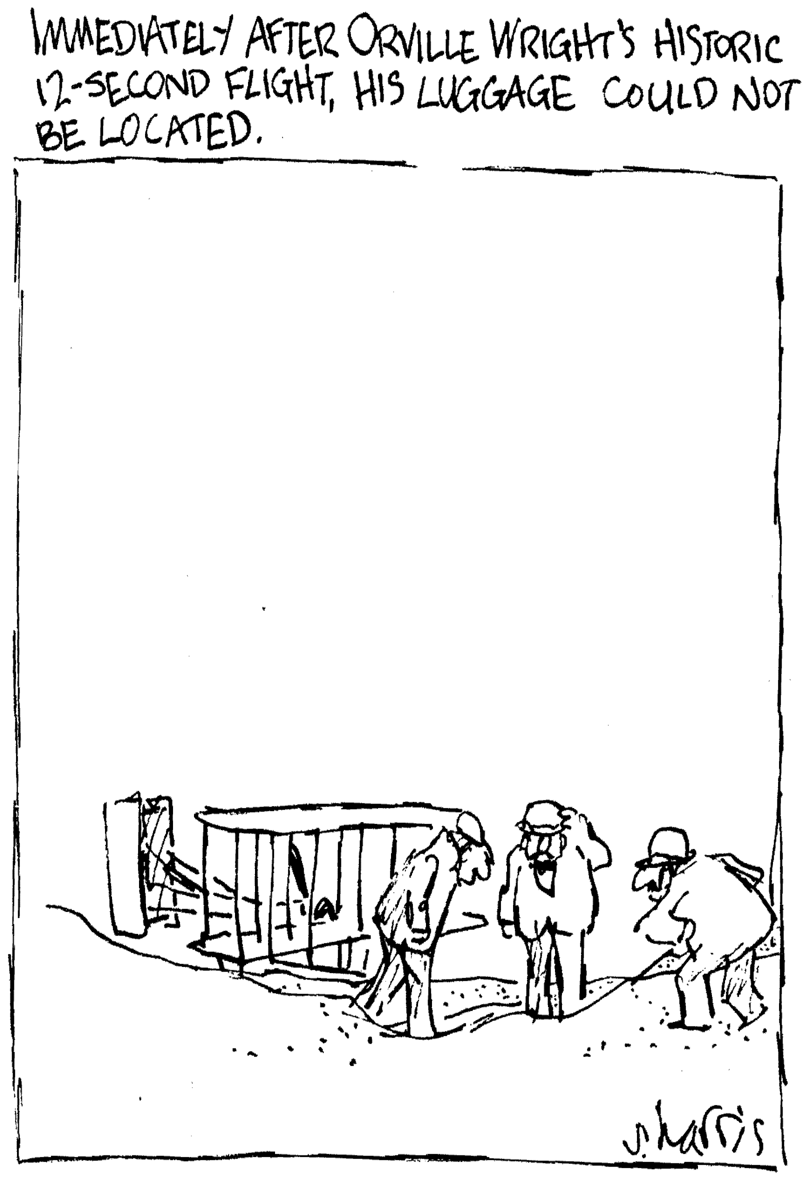


Figure 1.4: Early Traveling

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)* $\mathrm{Im}\, f = f(E)$, as the set

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

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

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

Proposition 1.8. *Given a linear map $f\colon E \to F$, the set $\mathrm{Im}\, f$ is a subspace of F and the set $\mathrm{Ker}\, f$ is a subspace of E. The linear map $f\colon E \to F$ is injective iff $\mathrm{Ker}\, f = 0$ (where 0 is the trivial subspace $\{0\}$).*

Since by Proposition 1.8, the image $\mathrm{Im}\, f$ of a linear map f is a subspace of F, we can define the *rank* $\mathrm{rk}(f)$ of f as the dimension of $\mathrm{Im}\, f$.

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 1.9. *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 1.9, 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 1.10. *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 1.9 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 1.9 is this:

Proposition 1.11. *Let E be a vector space of finite dimension $n \geq 1$ and let $f: E \to E$ 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 $\mathbf{GL}(n)$.

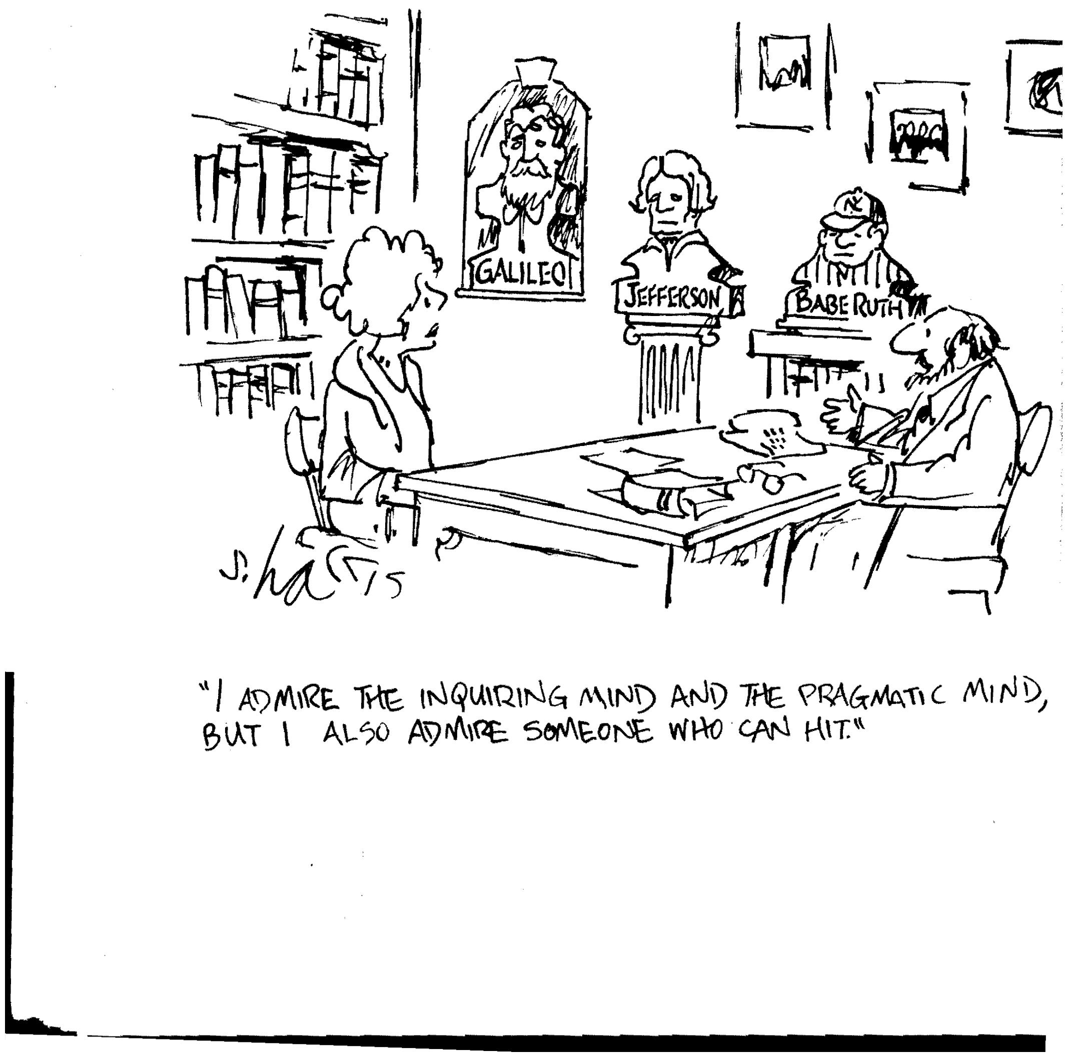


Figure 1.5: Hitting Power

1.6 Matrices

Proposition 1.9 shows that given two vector spaces E and F and a basis $(u_j)_{j\in J}$ of E, every linear map $f\colon E \to F$ is uniquely determined by the family $(f(u_j))_{j\in J}$ of the images under f of the vectors in the basis $(u_j)_{j\in J}$.

If we also have a basis $(v_i)_{i\in I}$ of F, then every vector $f(u_j)$ can be written in a unique way as

$$f(u_j) = \sum_{i\in I} a_{i\,j} v_i,$$

where $j \in J$, for a family of scalars $(a_{i\,j})_{i\in I}$.

Thus, with respect to the two bases $(u_j)_{j\in J}$ of E and $(v_i)_{i\in I}$ of F, the linear map f is completely determined by a "$I \times J$-matrix"

$$M(f) = (a_{i\,j})_{i\in I,\ j\in J}.$$

Remark: Note that we intentionally assigned the index set J to the basis $(u_j)_{j\in J}$ of E, and the index I to the basis $(v_i)_{i\in I}$ of F, so that the *rows* of the matrix $M(f)$ associated with $f\colon E \to F$ are indexed by I, and the *columns* of the matrix $M(f)$ are indexed by J.

Obviously, this causes a mildly unpleasant reversal. If we had considered the bases $(u_i)_{i\in I}$ of E and $(v_j)_{j\in J}$ of F, we would obtain a $J \times I$-matrix $M(f) = (a_{j\,i})_{j\in J,\ i\in I}$.

No matter what we do, there will be a reversal! We decided to stick to the bases $(u_j)_{j\in J}$ of E and $(v_i)_{i\in I}$ of F, so that we get an $I \times J$-matrix $M(f)$, knowing that we may occasionally suffer from this decision!

When I and J are finite, and say, when $|I| = m$ and $|J| = n$, the linear map f is determined by the matrix $M(f)$ whose entries in the j-th column are the components of the vector $f(u_j)$ over the basis $(v_1, \ldots, v_m)$, that is, the matrix

$$M(f) = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix}$$

whose entry on row i and column j is $a_{i\,j}$ $(1 \leq i \leq m$, $1 \leq j \leq n)$.

We will now show that when E and F have finite dimension, linear maps can be very conveniently represented by matrices, and that composition of linear maps corresponds to matrix multiplication.

We will follow rather closely an elegant presentation method due to Emil Artin.

Let E and F be two vector spaces, and assume that E has a finite basis $(u_1, \ldots, u_n)$ and that F has a finite basis $(v_1, \ldots, v_m)$. Recall that we have shown that every vector $x \in E$ can be written in a unique way as

$$x = x_1 u_1 + \cdots + x_n u_n,$$

and similarly every vector $y \in F$ can be written in a unique way as

$$y = y_1 v_1 + \cdots + y_m v_m.$$

Let $f \colon E \to F$ be a linear map between E and F.

Then, for every $x = x_1 u_1 + \cdots + x_n u_n$ in E, by linearity, we have

$$f(x) = x_1 f(u_1) + \cdots + x_n f(u_n).$$

Let

$$f(u_j) = a_{1\,j} v_1 + \cdots + a_{m\,j} v_m,$$

or more concisely,

$$f(u_j) = \sum_{i=1}^{m} a_{i\,j} v_i,$$

for every j, $1 \leq j \leq n$.

This can be expressed by writing the coefficients $a_{1j}, a_{2j}, \ldots, a_{mj}$ of $f(u_j)$ over the basis $(v_1, \ldots, v_m)$, as the jth column of a matrix, as shown below:

$$\begin{array}{c} \\ v_1 \\ v_2 \\ \vdots \\ v_m \end{array} \begin{array}{c} \begin{array}{cccc} f(u_1) & f(u_2) & \ldots & f(u_n) \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \end{array}$$

Then, substituting the right-hand side of each $f(u_j)$ into the expression for $f(x)$, we get

$$f(x) = x_1(\sum_{i=1}^{m} a_{i\,1}v_i) + \cdots + x_n(\sum_{i=1}^{m} a_{i\,n}v_i),$$

which, by regrouping terms to obtain a linear combination of the v_i, yields

$$f(x) = (\sum_{j=1}^{n} a_{1\,j}x_j)v_1 + \cdots + (\sum_{j=1}^{n} a_{m\,j}x_j)v_m.$$

Thus, letting $f(x) = y = y_1v_1 + \cdots + y_mv_m$, we have

$$y_i = \sum_{j=1}^{n} a_{i\,j}x_j \tag{1}$$

for all i, $1 \leq i \leq m$.

To make things more concrete, let us treat the case where $n = 3$ and $m = 2$.

In this case,

$$
\begin{aligned}
f(u_1) &= a_{11}v_1 + a_{21}v_2 \\
f(u_2) &= a_{12}v_1 + a_{22}v_2 \\
f(u_3) &= a_{13}v_1 + a_{23}v_2,
\end{aligned}
$$

which in matrix form is expressed by

$$
\begin{array}{c} \\ v_1 \\ v_2 \end{array}
\begin{array}{c} \begin{array}{ccc} f(u_1) & f(u_2) & f(u_3) \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \end{array},
$$

and for any $x = x_1u_1 + x_2u_2 + x_3u_3$, we have

$$
\begin{aligned}
f(x) &= f(x_1u_1 + x_2u_2 + x_3u_3) \\
&= x_1f(u_1) + x_2f(u_2) + x_3f(u_3) \\
&= x_1(a_{11}v_1 + a_{21}v_2) + x_2(a_{12}v_1 + a_{22}v_2) \\
&\quad + x_3(a_{13}v_1 + a_{23}v_2) \\
&= (a_{11}x_1 + a_{12}x_2 + a_{13}x_3)v_1 \\
&\quad + (a_{21}x_1 + a_{22}x_2 + a_{23}x_3)v_2.
\end{aligned}
$$

Consequently, since

$$
y = y_1v_1 + y_2v_2,
$$

we have

$$y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3$$
$$y_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3.$$

This agrees with the matrix equation

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Let us now consider how the composition of linear maps is expressed in terms of bases.

Let E, F, and G, be three vectors spaces with respective bases $(u_1, \ldots, u_p)$ for E, $(v_1, \ldots, v_n)$ for F, and $(w_1, \ldots, w_m)$ for G.

Let $g\colon E \to F$ and $f\colon F \to G$ be linear maps.

As explained earlier, $g\colon E \to F$ is determined by the images of the basis vectors u_j, and $f\colon F \to G$ is determined by the images of the basis vectors v_k.

We would like to understand how $f \circ g \colon E \to G$ is determined by the images of the basis vectors u_j.

Remark: Note that we are considering linear maps $g \colon E \to F$ and $f \colon F \to G$, instead of $f \colon E \to F$ and $g \colon F \to G$, which yields the composition $f \circ g \colon E \to G$ instead of $g \circ f \colon E \to G$.

Our perhaps unusual choice is motivated by the fact that if f is represented by a matrix $M(f) = (a_{i\,k})$ and g is represented by a matrix $M(g) = (b_{k\,j})$, then $f \circ g \colon E \to G$ is represented by the product AB of the matrices A and B.

If we had adopted the other choice where $f \colon E \to F$ and $g \colon F \to G$, then $g \circ f \colon E \to G$ would be represented by the product BA.

Obviously, this is a matter of taste! We will have to live with our perhaps unorthodox choice.

Thus, let

$$f(v_k) = \sum_{i=1}^{m} a_{i\,k} w_i,$$

for every k, $1 \leq k \leq n$, and let

$$g(u_j) = \sum_{k=1}^{n} b_{k\,j} v_k,$$

for every j, $1 \leq j \leq p$.

Also if

$$x = x_1 u_1 + \cdots + x_p u_p,$$

let

$$y = g(x)$$

and

$$z = f(y) = (f \circ g)(x),$$

with

$$y = y_1 v_1 + \cdots + y_n v_n$$

and

$$z = z_1 w_1 + \cdots + z_m w_m.$$

After some calculations, we get

$$z_i = \sum_{j=1}^{p} (\sum_{k=1}^{n} a_{i\,k} b_{k\,j}) x_j.$$

Thus, defining $c_{i\,j}$ such that

$$c_{i\,j} = \sum_{k=1}^{n} a_{i\,k} b_{k\,j},$$

for $1 \leq i \leq m$, and $1 \leq j \leq p$, we have

$$z_i = \sum_{j=1}^{p} c_{i\,j} x_j \tag{4}$$

Identity (4) suggests defining a multiplication operation on matrices, and we proceed to do so.

Definition 1.8. If $K = \mathbb{R}$ or $K = \mathbb{C}$, an *$m \times n$-matrix over K* is a family $(a_{i\,j})_{1 \leq i \leq m,\ 1 \leq j \leq n}$ of scalars in K, represented by an array

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix}$$

In the special case where $m = 1$, we have a *row vector*, represented by

$$(a_{1\,1} \cdots a_{1\,n})$$

and in the special case where $n = 1$, we have a *column vector*, represented by

$$\begin{pmatrix} a_{1\,1} \\ \vdots \\ a_{m\,1} \end{pmatrix}$$

In these last two cases, we usually omit the constant index 1 (first index in case of a row, second index in case of a column).

The set of all $m \times n$-matrices is denoted by $\mathrm{M}_{m,n}(K)$ or $\mathrm{M}_{m,n}$.

An $n \times n$-matrix is called a *square matrix of dimension n*.

The set of all square matrices of dimension n is denoted by $\mathrm{M}_n(K)$, or M_n.

Remark: As defined, a matrix $A = (a_{i\,j})_{1 \leq i \leq m,\ 1 \leq j \leq n}$ is a *family*, that is, a function from $\{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$ to K.

As such, there is no reason to assume an ordering on the indices. Thus, the matrix A can be represented in many different ways as an array, by adopting different orders for the rows or the columns.

However, it is customary (and usually convenient) to assume the natural ordering on the sets $\{1, 2, \ldots, m\}$ and $\{1, 2, \ldots, n\}$, and to represent A as an array according to this ordering of the rows and columns.

We also define some operations on matrices as follows.

Definition 1.9. Given two $m \times n$ matrices $A = (a_{i\,j})$ and $B = (b_{i\,j})$, we define their *sum* $A + B$ as the matrix $C = (c_{i\,j})$ such that $c_{i\,j} = a_{i\,j} + b_{i\,j}$; that is,

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix} + \begin{pmatrix} b_{1\,1} & b_{1\,2} & \ldots & b_{1\,n} \\ b_{2\,1} & b_{2\,2} & \ldots & b_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m\,1} & b_{m\,2} & \ldots & b_{m\,n} \end{pmatrix}$$
$$= \begin{pmatrix} a_{1\,1} + b_{1\,1} & a_{1\,2} + b_{1\,2} & \ldots & a_{1\,n} + b_{1\,n} \\ a_{2\,1} + b_{2\,1} & a_{2\,2} + b_{2\,2} & \ldots & a_{2\,n} + b_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} + b_{m\,1} & a_{m\,2} + b_{m\,2} & \ldots & a_{m\,n} + b_{m\,n} \end{pmatrix}.$$

We define the matrix $-A$ as the matrix $(-a_{i\,j})$.

Given a scalar $\lambda \in K$, we define the matrix λA as the matrix $C = (c_{i\,j})$ such that $c_{i\,j} = \lambda a_{i\,j}$; that is

$$\lambda \begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix} = \begin{pmatrix} \lambda a_{1\,1} & \lambda a_{1\,2} & \ldots & \lambda a_{1\,n} \\ \lambda a_{2\,1} & \lambda a_{2\,2} & \ldots & \lambda a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{m\,1} & \lambda a_{m\,2} & \ldots & \lambda a_{m\,n} \end{pmatrix}.$$

Given an $m \times n$ matrices $A = (a_{i\,k})$ and an $n \times p$ matrices $B = (b_{k\,j})$, we define their *product AB* as the $m \times p$ matrix $C = (c_{i\,j})$ such that

$$c_{i\,j} = \sum_{k=1}^{n} a_{i\,k} b_{k\,j},$$

for $1 \leq i \leq m$, and $1 \leq j \leq p$. In the product $AB = C$ shown below

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix} \begin{pmatrix} b_{1\,1} & b_{1\,2} & \ldots & b_{1\,p} \\ b_{2\,1} & b_{2\,2} & \ldots & b_{2\,p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n\,1} & b_{n\,2} & \ldots & b_{n\,p} \end{pmatrix} = \begin{pmatrix} c_{1\,1} & c_{1\,2} & \ldots & c_{1\,p} \\ c_{2\,1} & c_{2\,2} & \ldots & c_{2\,p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m\,1} & c_{m\,2} & \ldots & c_{m\,p} \end{pmatrix}$$

note that the entry of index i and j of the matrix AB obtained by multiplying the matrices A and B can be identified with the product of the *row matrix corresponding to the i-th row of A* with the *column matrix corresponding to the j-column of B*:

$$(a_{i\,1} \cdots a_{i\,n}) \begin{pmatrix} b_{1\,j} \\ \vdots \\ b_{n\,j} \end{pmatrix} = \sum_{k=1}^{n} a_{i\,k} b_{k\,j}.$$

The square matrix I_n of dimension n containing 1 on the diagonal and 0 everywhere else is called the *identity matrix*. It is denoted by

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Given an $m \times n$ matrix $A = (a_{i\,j})$, its *transpose* $A^\top = (a^\top_{j\,i})$, is the $n \times m$-matrix such that $a^\top_{j\,i} = a_{i\,j}$, for all i, $1 \leq i \leq m$, and all j, $1 \leq j \leq n$.

The transpose of a matrix A is sometimes denoted by A^t, or even by ${}^t A$.

Note that the transpose $A^\top$ of a matrix A has the property that the j-th row of $A^\top$ is the j-th column of A.

In other words, transposition exchanges the rows and the columns of a matrix.

The following observation will be useful later on when we discuss the SVD. Given any $m \times n$ matrix A and any $n \times p$ matrix B, if we denote the columns of A by $A^1, \ldots, A^n$ and the rows of B by $B_1, \ldots, B_n$, then we have

$$AB = A^1 B_1 + \cdots + A^n B_n.$$

For every square matrix A of dimension n, it is immediately verified that $AI_n = I_n A = A$.

If a matrix B such that $AB = BA = I_n$ exists, then it is unique, and it is called the *inverse* of A. The matrix B is also denoted by A^{-1}.

An invertible matrix is also called a *nonsingular* matrix, and a matrix that is not invertible is called a *singular* matrix.

Proposition 1.11 shows that if a square matrix A has a left inverse, that is a matrix B such that $BA = I$, or a right inverse, that is a matrix C such that $AC = I$, then A is actually invertible; so $B = A^{-1}$ and $C = A^{-1}$. This also follows from Proposition 1.25.

It is immediately verified that the set $\mathrm{M}_{m,n}(K)$ of $m \times n$ matrices is a *vector space* under addition of matrices and multiplication of a matrix by a scalar.

Consider the $m \times n$-matrices $E_{i,j} = (e_{h\,k})$, defined such that $e_{i\,j} = 1$, and $e_{h\,k} = 0$, if $h \neq i$ or $k \neq j$.

It is clear that every matrix $A = (a_{i\,j}) \in \mathrm{M}_{m,n}(K)$ can be written in a unique way as

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{i\,j} E_{i,j}.$$

Thus, the family $(E_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$ is a *basis* of the vector space $\mathrm{M}_{m,n}(K)$, which has dimension mn.

Square matrices provide a natural example of a noncommutative ring with zero divisors.

Example 1.8. For example, letting A, B be the 2×2-matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

then

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$BA = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

We now formalize the representation of linear maps by matrices.

Definition 1.10. Let E and F be two vector spaces, and let $(u_1, \ldots, u_n)$ be a basis for E, and $(v_1, \ldots, v_m)$ be a basis for F. Each vector $x \in E$ expressed in the basis $(u_1, \ldots, u_n)$ as $x = x_1 u_1 + \cdots + x_n u_n$ is represented by the column matrix

$$M(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and similarly for each vector $y \in F$ expressed in the basis $(v_1, \ldots, v_m)$. Every linear map $f\colon E \to F$ is represented by the matrix $M(f) = (a_{i\,j})$, where $a_{i\,j}$ is the i-th component of the vector $f(u_j)$ over the basis $(v_1, \ldots, v_m)$, i.e., where

$$f(u_j) = \sum_{i=1}^{m} a_{i\,j} v_i, \quad \text{for every } j,\ 1 \leq j \leq n.$$

The coefficients $a_{1j}, a_{2j}, \ldots, a_{mj}$ of $f(u_j)$ over the basis $(v_1, \ldots, v_m)$ form the jth column of the matrix $M(f)$ shown below:

$$\begin{array}{c} \\ v_1 \\ v_2 \\ \vdots \\ v_m \end{array} \begin{array}{c} \begin{array}{cccc} f(u_1) & f(u_2) & \ldots & f(u_n) \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \end{array}.$$

The matrix $M(f)$ associated with the linear map $f\colon E \to F$ is called the *matrix of f with respect to the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$*.

When $E = F$ and the basis $(v_1, \ldots, v_m)$ is identical to the basis $(u_1, \ldots, u_n)$ of E, the matrix $M(f)$ associated with $f\colon E \to E$ (as above) is called the *matrix of f with respect to the basis $(u_1, \ldots, u_n)$*.

Remark: As in the remark after Definition 1.8, there is no reason to assume that the vectors in the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ are ordered in any particular way.

However, it is often convenient to assume the natural ordering. When this is so, authors sometimes refer to the matrix $M(f)$ as the matrix of f with respect to the *ordered bases* $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$.

Then, given a linear map $f\colon E \to F$ represented by the matrix $M(f) = (a_{i\,j})$ w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$, by equations (1) and the definition of matrix multiplication, *the equation $y = f(x)$ corresponds to the matrix equation $M(y) = M(f)M(x)$*, that is,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{1\,1} & \ldots & a_{1\,n} \\ \vdots & \ddots & \vdots \\ a_{m\,1} & \ldots & a_{m\,n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Recall that

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= x_1 \begin{pmatrix} a_{1\,1} \\ a_{2\,1} \\ \vdots \\ a_{m\,1} \end{pmatrix} + x_2 \begin{pmatrix} a_{1\,2} \\ a_{2\,2} \\ \vdots \\ a_{m\,2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1\,n} \\ a_{2\,n} \\ \vdots \\ a_{m\,n} \end{pmatrix}.$$

Sometimes, it is necessary to incoporate the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ in the notation for the matrix $M(f)$ expressing f with respect to these bases. This turns out to be a messy enterprise!

We propose the following course of action: write $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{V} = (v_1, \ldots, v_m)$ for the bases of E and F, and denote by

$$M_{\mathcal{U},\mathcal{V}}(f)$$

the *matrix of f with respect to the bases $\mathcal{U}$ and $\mathcal{V}$*.

Furthermore, write $x_{\mathcal{U}}$ for the coordinates $M(x) = (x_1, \ldots, x_n)$ of $x \in E$ w.r.t. the basis $\mathcal{U}$ and write $y_{\mathcal{V}}$ for the coordinates $M(y) = (y_1, \ldots, y_m)$ of $y \in F$ w.r.t. the basis $\mathcal{V}$. Then,

$$y = f(x)$$

is expressed in matrix form by

$$y_{\mathcal{V}} = M_{\mathcal{U},\mathcal{V}}(f)\, x_{\mathcal{U}}.$$

When $\mathcal{U} = \mathcal{V}$, we abbreviate $M_{\mathcal{U},\mathcal{V}}(f)$ as $M_{\mathcal{U}}(f)$.

The above notation seems reasonable, but it has the slight disadvantage that in the expression $M_{\mathcal{U},\mathcal{V}}(f)x_{\mathcal{U}}$, the input argument $x_{\mathcal{U}}$ which is fed to the matrix $M_{\mathcal{U},\mathcal{V}}(f)$ does not appear next to the subscript $\mathcal{U}$ in $M_{\mathcal{U},\mathcal{V}}(f)$.

We could have used the notation $M_{\mathcal{V},\mathcal{U}}(f)$, and some people do that. But then, we find a bit confusing that $\mathcal{V}$ comes before $\mathcal{U}$ when f maps from the space E with the basis $\mathcal{U}$ to the space F with the basis $\mathcal{V}$.

So, we prefer to use the notation $M_{\mathcal{U},\mathcal{V}}(f)$.

Be aware that other authors such as Meyer [25] use the notation $[f]_{\mathcal{U},\mathcal{V}}$, and others such as Dummit and Foote [13] use the notation $M_{\mathcal{U}}^{\mathcal{V}}(f)$, instead of $M_{\mathcal{U},\mathcal{V}}(f)$.

This gets worse! You may find the notation $M_{\mathcal{V}}^{\mathcal{U}}(f)$ (as in Lang [21]), or ${}_{\mathcal{U}}[f]_{\mathcal{V}}$, or other strange notations.

Let us illustrate the representation of a linear map by a matrix in a concrete situation.

Let E be the vector space $\mathbb{R}[X]_4$ of polynomials of degree at most 4, let F be the vector space $\mathbb{R}[X]_3$ of polynomials of degree at most 3, and let the linear map be the derivative map d: that is,

$$
\begin{aligned}
d(P+Q) &= dP + dQ \\
d(\lambda P) &= \lambda dP,
\end{aligned}
$$

with $\lambda \in \mathbb{R}$.

We choose $(1, x, x^2, x^3, x^4)$ as a basis of E and $(1, x, x^2, x^3)$ as a basis of F.

Then, the 4×5 matrix D associated with d is obtained by expressing the derivative dx^i of each basis vector x^i for $i = 0, 1, 2, 3, 4$ over the basis $(1, x, x^2, x^3)$.

We find

$$D = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

Then, if P denotes the polynomial

$$P = 3x^4 - 5x^3 + x^2 - 7x + 5,$$

we have

$$dP = 12x^3 - 15x^2 + 2x - 7,$$

the polynomial P is represented by the vector $(5, -7, 1, -5, 3)$ and dP is represented by the vector $(-7, 2, -15, 12)$, and we have

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 5 \\ -7 \\ 1 \\ -5 \\ 3 \end{pmatrix} = \begin{pmatrix} -7 \\ 2 \\ -15 \\ 12 \end{pmatrix},$$

as expected!

The kernel (nullspace) of d consists of the polynomials of degree 0, that is, the constant polynomials.

Therefore $\dim(\mathrm{Ker}\, d) = 1$, and from

$$\dim(E) = \dim(\mathrm{Ker}\, d) + \dim(\mathrm{Im}\, d)$$

(see Theorem 1.22), we get $\dim(\mathrm{Im}\, d) = 4$
(since $\dim(E) = 5$).

For fun, let us figure out the linear map from the vector space $\mathbb{R}[X]_3$ to the vector space $\mathbb{R}[X]_4$ given by integration (finding the primitive, or anti-derivative) of x^i, for $i = 0, 1, 2, 3$).

The 5×4 matrix S representing $\int$ with respect to the same bases as before is

$$S = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix}.$$

We verify that $DS = I_4$,

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

as it should!

The equation $DS = I_4$ show that S is injective and has D as a left inverse. However, $SD \neq I_5$, and instead

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

because constant polynomials (polynomials of degree 0) belong to the kernel of D.

The function that associates to a linear map $f\colon E \to F$ the matrix $M(f)$ w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ has the property that matrix multiplication corresponds to composition of linear maps.

This allows us to transfer properties of linear maps to matrices.

Proposition 1.12. *(1) Given any matrices* $A \in \mathrm{M}_{m,n}(K)$, $B \in \mathrm{M}_{n,p}(K)$, *and* $C \in \mathrm{M}_{p,q}(K)$, *we have*

$$(AB)C = A(BC);$$

that is, matrix multiplication is associative.

(2) Given any matrices $A, B \in \mathrm{M}_{m,n}(K)$, *and* $C, D \in \mathrm{M}_{n,p}(K)$, *for all* $\lambda \in K$, *we have*

$$\begin{aligned}(A + B)C &= AC + BC \\ A(C + D) &= AC + AD \\ (\lambda A)C &= \lambda(AC) \\ A(\lambda C) &= \lambda(AC),\end{aligned}$$

so that matrix multiplication $\cdot\colon \mathrm{M}_{m,n}(K) \times \mathrm{M}_{n,p}(K) \to \mathrm{M}_{m,p}(K)$ *is bilinear.*

Note that Proposition 1.12 implies that the vector space $\mathrm{M}_n(K)$ of square matrices is a (noncommutative) *ring* with unit I_n.

The following proposition states the main properties of the mapping $f \mapsto M(f)$ between $\mathrm{Hom}(E, F)$ and $\mathrm{M}_{m,n}$.

In short, it is an isomorphism of vector spaces.

Proposition 1.13. *Given three vector spaces E, F, G, with respective bases $(u_1, \ldots, u_p)$, $(v_1, \ldots, v_n)$, and $(w_1, \ldots, w_m)$, the mapping $M \colon \mathrm{Hom}(E, F) \to \mathrm{M}_{n,p}$ that associates the matrix $M(g)$ to a linear map $g \colon E \to F$ satisfies the following properties for all $x \in E$, all $g, h \colon E \to F$, and all $f \colon F \to G$:*

$$\begin{aligned} M(g(x)) &= M(g)M(x) \\ M(g + h) &= M(g) + M(h) \\ M(\lambda g) &= \lambda M(g) \\ M(f \circ g) &= M(f)M(g). \end{aligned}$$

Thus, $M \colon \mathrm{Hom}(E, F) \to \mathrm{M}_{n,p}$ is an isomorphism of vector spaces, and when $p = n$ and the basis $(v_1, \ldots, v_n)$ is identical to the basis $(u_1, \ldots, u_p)$,
$M \colon \mathrm{Hom}(E, E) \to \mathrm{M}_n$ is an isomorphism of rings.

In view of Proposition 1.13, it seems preferable to represent vectors from a vector space of finite dimension as column vectors rather than row vectors.

Thus, from now on, we will denote vectors of $\mathbb{R}^n$ (or more generally, of K^n) as columm vectors.

It is important to observe that the isomorphism $M\colon \mathrm{Hom}(E, F) \to \mathrm{M}_{n,p}$ given by Proposition 1.13 depends on the choice of the bases $(u_1, \ldots, u_p)$ and $(v_1, \ldots, v_n)$, and similarly for the isomorphism $M\colon \mathrm{Hom}(E, E) \to \mathrm{M}_n$, which depends on the choice of the basis $(u_1, \ldots, u_n)$.

Thus, it would be useful to know how a change of basis affects the representation of a linear map $f\colon E \to F$ as a matrix.

Proposition 1.14. *Let E be a vector space, and let $(u_1, \ldots, u_n)$ be a basis of E. For every family $(v_1, \ldots, v_n)$, let $P = (a_{i\,j})$ be the matrix defined such that $v_j = \sum_{i=1}^n a_{i\,j} u_i$. The matrix P is invertible iff $(v_1, \ldots, v_n)$ is a basis of E.*

Definition 1.11. Given a vector space E of dimension n, for any two bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ of E, let $P = (a_{i\,j})$ be the invertible matrix defined such that

$$v_j = \sum_{i=1}^n a_{i\,j} u_i,$$

which is also the matrix of the identity $\mathrm{id}\colon E \to E$ with respect to the bases $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_n)$, *in that order*. Indeed, we express each $\mathrm{id}(v_j) = v_j$ over the basis $(u_1, \ldots, u_n)$. The coefficients $a_{1j}, a_{2j}, \ldots, a_{nj}$ of v_j over the basis $(u_1, \ldots, u_n)$ form the jth column of the matrix P shown below:

$$\begin{array}{c} \\ u_1 \\ u_2 \\ \vdots \\ u_n \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & \ldots & v_n \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix} \end{array}.$$

The matrix P is called the *change of basis matrix from* $(u_1, \ldots, u_n)$ *to* $(v_1, \ldots, v_n)$.

Clearly, the change of basis matrix from $(v_1, \ldots, v_n)$ to $(u_1, \ldots, u_n)$ is P^{-1}.

Since $P = (a_{i\,j})$ is the matrix of the identity $\mathrm{id}\colon E \to E$ with respect to the bases $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_n)$, given any vector $x \in E$, if $x = x_1 u_1 + \cdots + x_n u_n$ over the basis $(u_1, \ldots, u_n)$ and $x = x'_1 v_1 + \cdots + x'_n v_n$ over the basis $(v_1, \ldots, v_n)$, from Proposition 1.13, we have

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{1\,1} & \ldots & a_{1\,n} \\ \vdots & \ddots & \vdots \\ a_{n\,1} & \ldots & a_{n\,n} \end{pmatrix} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

showing that the *old* coordinates (x_i) of x (over $(u_1, \ldots, u_n)$) are expressed in terms of the *new* coordinates (x'_i) of x (over $(v_1, \ldots, v_n)$).

Now we face the painful task of assigning a “good” notation incorporating the bases $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{V} = (v_1, \ldots, v_n)$ into the notation for the change of basis matrix from $\mathcal{U}$ to $\mathcal{V}$.

Because the change of basis matrix from $\mathcal{U}$ to $\mathcal{V}$ is the matrix of the identity map id_E *with respect to the bases* $\mathcal{V}$ *and* $\mathcal{U}$ *in that order*, we could denote it by $M_{\mathcal{V},\mathcal{U}}(\mathrm{id})$ (Meyer [25] uses the notation $[I]_{\mathcal{V},\mathcal{U}}$).

We prefer to use an abbreviation for $M_{\mathcal{V},\mathcal{U}}(\mathrm{id})$ and we will use the notation

$$P_{\mathcal{V},\mathcal{U}}$$

for the *change of basis matrix from* $\mathcal{U}$ *to* $\mathcal{V}$.

Note that

$$P_{\mathcal{U},\mathcal{V}} = P_{\mathcal{V},\mathcal{U}}^{-1}.$$

Then, if we write $x_{\mathcal{U}} = (x_1, \ldots, x_n)$ for the *old* coordinates of x with respect to the basis $\mathcal{U}$ and $x_{\mathcal{V}} = (x'_1, \ldots, x'_n)$ for the *new* coordinates of x with respect to the basis $\mathcal{V}$, we have

$$x_{\mathcal{U}} = P_{\mathcal{V},\mathcal{U}}\, x_{\mathcal{V}}, \quad x_{\mathcal{V}} = P_{\mathcal{V},\mathcal{U}}^{-1}\, x_{\mathcal{U}}.$$

The above may look backward, but remember that the matrix $M_{\mathcal{U},\mathcal{V}}(f)$ takes input expressed over the basis $\mathcal{U}$ to output expressed over the basis $\mathcal{V}$.

Consequently, $P_{\mathcal{V},\mathcal{U}}$ takes input expressed over the basis $\mathcal{V}$ to output expressed over the basis $\mathcal{U}$, and $x_{\mathcal{U}} = P_{\mathcal{V},\mathcal{U}}\, x_{\mathcal{V}}$ matches this point of view!

Beware that some authors (such as Artin [1]) define the change of basis matrix from $\mathcal{U}$ to $\mathcal{V}$ as $P_{\mathcal{U},\mathcal{V}} = P_{\mathcal{V},\mathcal{U}}^{-1}$. Under this point of view, the old basis $\mathcal{U}$ is expressed in terms of the new basis $\mathcal{V}$. We find this a bit unnatural.

Also, in practice, it seems that the new basis is often expressed in terms of the old basis, rather than the other way around.

Since the matrix $P = P_{\mathcal{V},\mathcal{U}}$ expresses the *new* basis $(v_1, \ldots, v_n)$ in terms of the *old* basis $(u_1, \ldots, u_n)$, we observe that the coordinates (x_i) of a vector x vary in the *opposite direction* of the change of basis.

For this reason, vectors are sometimes said to be *contravariant*.

However, this expression does not make sense! Indeed, a vector in an intrinsic quantity that does not depend on a specific basis.

What makes sense is that the *coordinates* of a vector vary in a contravariant fashion.

Let us consider some concrete examples of change of bases.

Example 1.9. Let $E = F = \mathbb{R}^2$, with $u_1 = (1, 0)$, $u_2 = (0, 1)$, $v_1 = (1, 1)$ and $v_2 = (-1, 1)$.

The change of basis matrix P from the basis $\mathcal{U} = (u_1, u_2)$ to the basis $\mathcal{V} = (v_1, v_2)$ is

$$P = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

and its inverse is

$$P^{-1} = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$

The old coordinates (x_1, x_2) with respect to (u_1, u_2) are expressed in terms of the new coordinates (x_1', x_2') with respect to (v_1, v_2) by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \end{pmatrix},$$

and the new coordinates (x_1', x_2') with respect to (v_1, v_2) are expressed in terms of the old coordinates (x_1, x_2) with respect to (u_1, u_2) by

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Example 1.10. Let $E = F = \mathbb{R}[X]_3$ be the set of polynomials of degree at most 3, and consider the bases $\mathcal{U} = (1, x, x^2, x^3)$ and $\mathcal{V} = (B_0^3(x), B_1^3(x), B_2^3(x), B_3^3(x))$, where $B_0^3(x), B_1^3(x), B_2^3(x), B_3^3(x)$ are the *Bernstein polynomials* of degree 3, given by

$$\begin{aligned} B_0^3(x) &= (1-x)^3 & B_1^3(x) &= 3(1-x)^2 x \\ B_2^3(x) &= 3(1-x)x^2 & B_3^3(x) &= x^3. \end{aligned}$$

By expanding the Bernstein polynomials, we find that the change of basis matrix $P_{\mathcal{V},\mathcal{U}}$ is given by

$$P_{\mathcal{V},\mathcal{U}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{pmatrix}.$$

We also find that the inverse of $P_{\mathcal{V},\mathcal{U}}$ is

$$P_{\mathcal{V},\mathcal{U}}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1/3 & 0 & 0 \\ 1 & 2/3 & 1/3 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Therefore, the coordinates of the polynomial $2x^3 - x + 1$ over the basis $\mathcal{V}$ are

$$\begin{pmatrix} 1 \\ 2/3 \\ 1/3 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1/3 & 0 & 0 \\ 1 & 2/3 & 1/3 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix},$$

and so

$$2x^3 - x + 1 = B_0^3(x) + \frac{2}{3}B_1^3(x) + \frac{1}{3}B_2^3(x) + 2B_3^3(x).$$

Our next example is the Haar wavelets, a fundamental tool in signal processing.

1.7 Haar Basis Vectors and a Glimpse at Wavelets

We begin by considering *Haar wavelets* in $\mathbb{R}^4$.

Wavelets play an important role in audio and video signal processing, especially for *compressing* long signals into much smaller ones than still retain enough information so that when they are played, we can't see or hear any difference.

Consider the four vectors w_1, w_2, w_3, w_4 given by

$$w_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad w_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \quad w_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} \quad w_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

Note that these vectors are pairwise orthogonal, so they are indeed linearly independent (we will see this in a later chapter).

Let $\mathcal{W} = \{w_1, w_2, w_3, w_4\}$ be the *Haar basis*, and let $\mathcal{U} = \{e_1, e_2, e_3, e_4\}$ be the canonical basis of $\mathbb{R}^4$.

The change of basis matrix $W = P_{\mathcal{W},\mathcal{U}}$ from $\mathcal{U}$ to $\mathcal{W}$ is given by

$$W = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{pmatrix},$$

and we easily find that the inverse of W is given by

$$W^{-1} = \begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

So, the vector $v = (6, 4, 5, 1)$ over the basis $\mathcal{U}$ becomes $c = (c_1, c_2, c_3, c_4) = (4, 1, 1, 2)$ over the Haar basis $\mathcal{W}$, with

$$\begin{pmatrix} 4 \\ 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 6 \\ 4 \\ 5 \\ 1 \end{pmatrix}.$$

Given a signal $v = (v_1, v_2, v_3, v_4)$, we first *transform* v into its coefficients $c = (c_1, c_2, c_3, c_4)$ over the Haar basis by computing $c = W^{-1}v$. Observe that

$$c_1 = \frac{v_1 + v_2 + v_3 + v_4}{4}$$

is the overall *average* value of the signal v. The coefficient c_1 corresponds to the background of the image (or of the sound).

Then, c_2 gives the coarse details of v, whereas, c_3 gives the details in the first part of v, and c_4 gives the details in the second half of v.

Reconstruction of the signal consists in computing $v = Wc$.

The trick for good *compression* is to throw away some of the coefficients of c (set them to zero), obtaining a *compressed signal* $\widehat{c}$, and still retain enough crucial information so that the reconstructed signal $\widehat{v} = W\widehat{c}$ looks almost as good as the original signal v.

Thus, the steps are:

$$\text{input}\, v \;\longrightarrow\; \text{coefficients}\, c = W^{-1}v \;\longrightarrow\; \text{compressed}\, \widehat{c} \;\longrightarrow\; \text{compressed}\; \widehat{v} = W\widehat{c}.$$

This kind of compression scheme makes modern video conferencing possible.

It turns out that there is a faster way to find $c = W^{-1}v$, without actually using W^{-1}. This has to do with the multiscale nature of Haar wavelets.

Given the original signal $v = (6, 4, 5, 1)$ shown in Figure 1.6, we compute averages and half differences obtaining

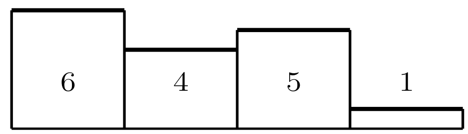


Figure 1.6: The original signal v

Figure 1.7: We get the coefficients $c_3 = 1$ and $c_4 = 2$.

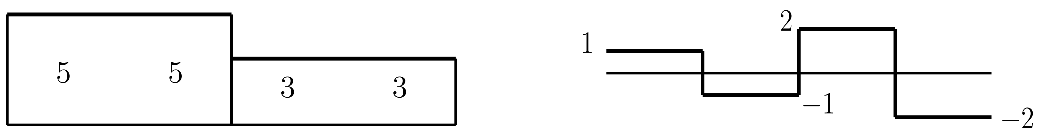


Figure 1.7: First averages and first half differences

Note that the original signal v can be reconstruced from the two signals in Figure 1.7.

Then, again we compute averages and half differences obtaining Figure 1.8.

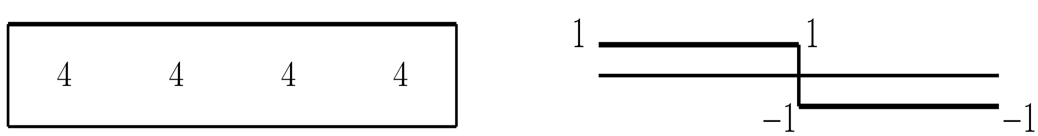


Figure 1.8: Second averages and second half differences

We get the coefficients $c_1 = 4$ and $c_2 = 1$.

Again, the signal on the left of Figure 1.7 can be reconstructed from the two signals in Figure 1.8.

This method can be generalized to signals of any length 2^n. The previous case corresponds to $n = 2$.

Let us consider the case $n = 3$.

The Haar basis $(w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8)$ is given by the matrix

$$W = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & -1 \end{pmatrix}$$

The columns of this matrix are orthogonal and it is easy to see that

$$W^{-1} = \mathrm{diag}(1/8, 1/8, 1/4, 1/4, 1/2, 1/2, 1/2, 1/2)W^{\top}.$$

A pattern is begining to emerge. It looks like the second Haar basis vector w_2 is the "mother" of all the other basis vectors, except the first, whose purpose is to perform averaging.

Indeed, in general, given

$$w_2 = \underbrace{(1, \ldots, 1, -1, \ldots, -1)}_{2^n},$$

the other Haar basis vectors are obtained by a "scaling and shifting process."

Starting from w_2, the scaling process generates the vectors

$$w_3, w_5, w_9, \ldots, w_{2^j+1}, \ldots, w_{2^{n-1}+1},$$

such that $w_{2^{j+1}+1}$ is obtained from w_{2^j+1} by forming two consecutive blocks of 1 and -1 of half the size of the blocks in w_{2^j+1}, and setting all other entries to zero. Observe that w_{2^j+1} has 2^j blocks of 2^{n-j} elements.

The shifting process, consists in shifting the blocks of 1 and -1 in w_{2^j+1} to the right by inserting a block of $(k-1)2^{n-j}$ zeros from the left, with $0 \leq j \leq n-1$ and $1 \leq k \leq 2^j$.

Thus, we obtain the following formula for w_{2^j+k}:

$$w_{2^j+k}(i) = \begin{cases} 0 & 1 \leq i \leq (k-1)2^{n-j} \\ 1 & (k-1)2^{n-j}+1 \leq i \leq (k-1)2^{n-j}+2^{n-j-1} \\ -1 & (k-1)2^{n-j}+2^{n-j-1}+1 \leq i \leq k2^{n-j} \\ 0 & k2^{n-j}+1 \leq i \leq 2^n, \end{cases}$$

with $0 \leq j \leq n-1$ and $1 \leq k \leq 2^j$.

Of course

$$w_1 = \underbrace{(1, \ldots, 1)}_{2^n}.$$

The above formulae look a little better if we change our indexing slightly by letting k vary from 0 to $2^j - 1$ and using the index j instead of 2^j.

In this case, the Haar basis is denoted by

$$w_1, h_0^0, h_0^1, h_1^1, h_0^2, h_1^2, h_2^2, h_3^2, \ldots, h_k^j, \ldots, h_{2^{n-1}-1}^{n-1},$$

and

$$h_k^j(i) = \begin{cases} 0 & 1 \leq i \leq k2^{n-j} \\ 1 & k2^{n-j} + 1 \leq i \leq k2^{n-j} + 2^{n-j-1} \\ -1 & k2^{n-j} + 2^{n-j-1} + 1 \leq i \leq (k+1)2^{n-j} \\ 0 & (k+1)2^{n-j} + 1 \leq i \leq 2^n, \end{cases}$$

with $0 \leq j \leq n-1$ and $0 \leq k \leq 2^j - 1$.

It turns out that there is a way to understand these formulae better if we interpret a vector $u = (u_1, \ldots, u_m)$ as a piecewise linear function over the interval $[0, 1)$.

We define the function $\mathrm{plf}(u)$ such that

$$\mathrm{plf}(u)(x) = u_i, \qquad \frac{i-1}{m} \leq x < \frac{i}{m},\ 1 \leq i \leq m.$$

In words, the function $\mathrm{plf}(u)$ has the value u_1 on the interval $[0, 1/m)$, the value u_2 on $[1/m, 2/m)$, etc., and the value u_m on the interval $[(m-1)/m, 1)$.

For example, the piecewise linear function associated with the vector

$$u = (2.4, 2.2, 2.15, 2.05, 6.8, 2.8, -1.1, -1.3)$$

is shown in Figure 1.9.

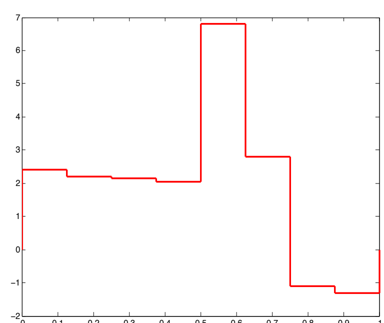


Figure 1.9: The piecewise linear function $\mathrm{plf}(u)$

Then, each basis vector h^j_k corresponds to the function

$$\psi^j_k = \mathrm{plf}(h^j_k).$$

In particular, for all n, the Haar basis vectors

$$h^0_0 = w_2 = \underbrace{(1, \ldots, 1, -1, \ldots, -1)}_{2^n}$$

yield the same piecewise linear function ψ given by

$$\psi(x) = \begin{cases} 1 & \text{if} \quad 0 \leq x < 1/2 \\ -1 & \text{if} \quad 1/2 \leq x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

whose graph is shown in Figure 1.10.

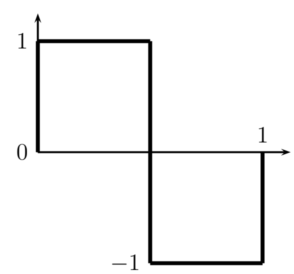


Figure 1.10: The Haar wavelet ψ

Then, it is easy to see that ψ_k^j is given by the simple expression

$$\psi_k^j(x) = \psi(2^j x - k), \quad 0 \leq j \leq n-1,\ 0 \leq k \leq 2^j - 1.$$

The above formula makes it clear that ψ_k^j is obtained from ψ by scaling and shifting.

The function $\phi_0^0 = \mathrm{plf}(w_1)$ is the piecewise linear function with the constant value 1 on $[0,1)$, and the functions ψ_k^j together with φ_0^0 are known as the *Haar wavelets*.

Rather than using W^{-1} to convert a vector u to a vector c of coefficients over the Haar basis, and the matrix W to reconstruct the vector u from its Haar coefficients c, we can use faster algorithms that use averaging and differencing.

If c is a vector of Haar coefficients of dimension 2^n, we compute the sequence of vectors $u_0, u_1, \ldots, u_n$ as follows:

$$\begin{aligned} u_0 &= c \\ u_{j+1} &= u_j \\ u_{j+1}(2i-1) &= u_j(i) + u_j(2^j + i) \\ u_{j+1}(2i) &= u_j(i) - u_j(2^j + i), \end{aligned}$$

for $j = 0, \ldots, n-1$ and $i = 1, \ldots, 2^j$.

The reconstructed vector (signal) is $u = u_n$.

If u is a vector of dimension 2^n, we compute the sequence of vectors $c_n, c_{n-1}, \ldots, c_0$ as follows:

$$\begin{aligned} c_n &= u \\ c_j &= c_{j+1} \\ c_j(i) &= (c_{j+1}(2i-1) + c_{j+1}(2i))/2 \\ c_j(2^j + i) &= (c_{j+1}(2i-1) - c_{j+1}(2i))/2, \end{aligned}$$

for $j = n-1, \ldots, 0$ and $i = 1, \ldots, 2^j$.

The vector over the Haar basis is $c = c_0$.

Here is an example of the conversion of a vector to its Haar coefficients for $n = 3$.

Given the sequence $u = (31, 29, 23, 17, -6, -8, -2, -4)$, we get the sequence

$$\begin{aligned}
c_3 &= (31, 29, 23, 17, -6, -8, -2, -4)\\
c_2 &= (30, 20, -7, -3, 1, 3, 1, 1)\\
c_1 &= (25, -5, 5, -2, 1, 3, 1, 1)\\
c_0 &= (10, 15, 5, -2, 1, 3, 1, 1),
\end{aligned}$$

so $c = (10, 15, 5, -2, 1, 3, 1, 1)$.

Conversely, given $c = (10, 15, 5, -2, 1, 3, 1, 1)$, we get the sequence

$$\begin{aligned}
u_0 &= (10, 15, 5, -2, 1, 3, 1, 1)\\
u_1 &= (25, -5, 5, -2, 1, 3, 1, 1)\\
u_2 &= (30, 20, -7, -3, 1, 3, 1, 1)\\
u_3 &= (31, 29, 23, 17, -6, -8, -2, -4),
\end{aligned}$$

which gives back $u = (31, 29, 23, 17, -6, -8, -2, -4)$.

An important and attractive feature of the Haar basis is that it provides a *multiresolution analysis* of a signal.

Indeed, given a signal u, if $c = (c_1, \ldots, c_{2^n})$ is the vector of its Haar coefficients, the coefficients with low index give coarse information about u, and the coefficients with high index represent fine information.

This multiresolution feature of wavelets can be exploited to compress a signal, that is, to use fewer coefficients to represent it. Here is an example.

Consider the signal

$$u = (2.4, 2.2, 2.15, 2.05, 6.8, 2.8, -1.1, -1.3),$$

whose Haar transform is

$$c = (2, 0.2, 0.1, 3, 0.1, 0.05, 2, 0.1).$$

The piecewise-linear curves corresponding to u and c are shown in Figure 1.11.

Since some of the coefficients in c are small (smaller than or equal to 0.2) we can compress c by replacing them by 0.

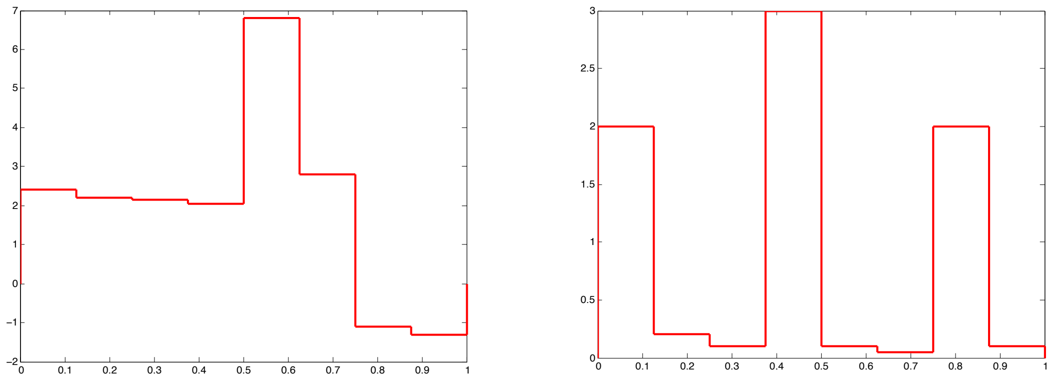


Figure 1.11: A signal and its Haar transform

We get

$$c_2 = (2, 0, 0, 3, 0, 0, 2, 0),$$

and the reconstructed signal is

$$u_2 = (2, 2, 2, 2, 7, 3, -1, -1).$$

The piecewise-linear curves corresponding to u_2 and c_2 are shown in Figure 1.12.

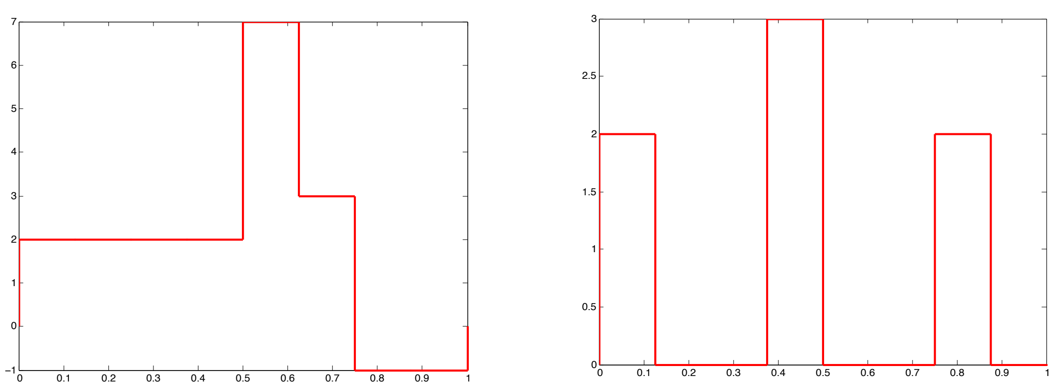


Figure 1.12: A compressed signal and its compressed Haar transform

An interesting (and amusing) application of the Haar wavelets is to the compression of audio signals.

It turns out that if your type `load handel` in `Matlab` an audio file will be loaded in a vector denoted by y, and if you type **`sound(y)`**, the computer will play this piece of music.

You can convert y to its vector of Haar coefficients, c. The length of y is 73113, so first tuncate the tail of y to get a vector of length $65536 = 2^{16}$.

A plot of the signals corresponding to y and c is shown in Figure 1.13.

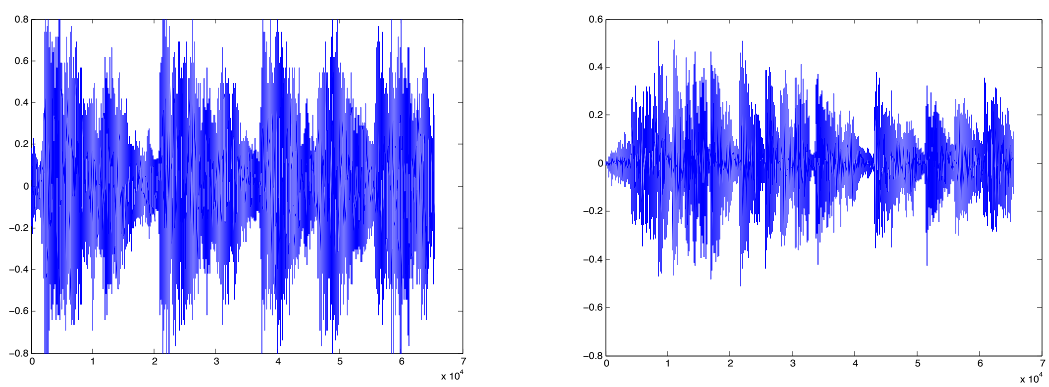


Figure 1.13: The signal "handel" and its Haar transform

Then, run a program that sets all coefficients of c whose absolute value is less that 0.05 to zero. This sets 37272 coefficients to 0.

The resulting vector c_2 is converted to a signal y_2. A plot of the signals corresponding to y_2 and c_2 is shown in Figure 1.14.

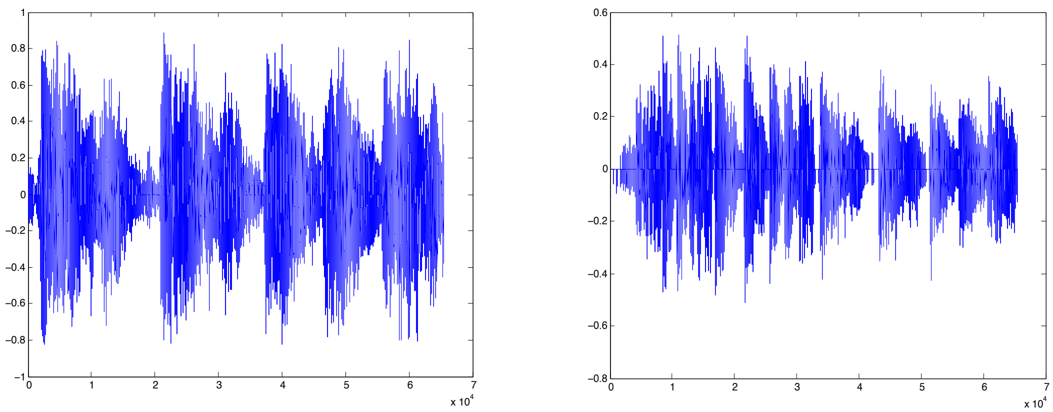


Figure 1.14: The compressed signal "handel" and its Haar transform

When you type **`sound(y2)`**, you find that the music doesn't differ much from the original, although it sounds less crisp.

Another neat property of the Haar transform is that it can be instantly generalized to matrices (even rectangular) without any extra effort!

This allows for the compression of digital images. But first, we address the issue of normalization of the Haar coefficients.

As we observed earlier, the $2^n \times 2^n$ matrix W_n of Haar basis vectors has orthogonal columns, but its columns do not have unit length.

As a consequence, $W_n^\top$ is not the inverse of W_n, but rather the matrix

$$W_n^{-1} = D_n W_n^\top$$

with

$$\begin{aligned} D_n = \operatorname{diag}\Big(2^{-n}, \underbrace{2^{-n}}_{2^0}, \underbrace{2^{-(n-1)}, 2^{-(n-1)}}_{2^1}, \\ \underbrace{2^{-(n-2)}, \ldots, 2^{-(n-2)}}_{2^2}, \ldots, \underbrace{2^{-1}, \ldots, 2^{-1}}_{2^{n-1}}\Big). \end{aligned}$$

Therefore, we define the orthogonal matrix

$$H_n = W_n D_n^{\frac{1}{2}}$$

whose columns are the normalized Haar basis vectors, with

$$D_n^{\frac{1}{2}} = \mathrm{diag}\Big(2^{-\frac{n}{2}}, \underbrace{2^{-\frac{n}{2}}}_{2^0}, \underbrace{2^{-\frac{n-1}{2}}, 2^{-\frac{n-1}{2}}}_{2^1}, \underbrace{2^{-\frac{n-2}{2}}, \ldots, 2^{-\frac{n-2}{2}}}_{2^2}, \ldots, \underbrace{2^{-\frac{1}{2}}, \ldots, 2^{-\frac{1}{2}}}_{2^{n-1}}\Big).$$

We call H_n the *normalized Haar transform matrix*.

Because H_n is orthogonal, $H_n^{-1} = H_n^{\top}$.

Given a vector (signal) u, we call $c = H_n^{\top} u$ the *normalized Haar coefficients* of u.

When computing the sequence of u_js, use

$$\begin{aligned} u_{j+1}(2i-1) &= (u_j(i) + u_j(2^j + i))/\sqrt{2} \\ u_{j+1}(2i) &= (u_j(i) - u_j(2^j + i))/\sqrt{2}, \end{aligned}$$

and when computing the sequence of c_js, use

$$\begin{aligned} c_j(i) &= (c_{j+1}(2i-1) + c_{j+1}(2i))/\sqrt{2} \\ c_j(2^j + i) &= (c_{j+1}(2i-1) - c_{j+1}(2i))/\sqrt{2}. \end{aligned}$$

Note that things are now more symmetric, at the expense of a division by $\sqrt{2}$. However, for long vectors, it turns out that these algorithms are numerically more stable.

Let us now explain the 2D version of the Haar transform.

We describe the version using the matrix W_n, the method using H_n being identical (except that $H_n^{-1} = H_n^\top$, but this does not hold for W_n^{-1}).

Given a $2^m \times 2^n$ matrix A, we can first convert the *rows* of A to their Haar coefficients using the Haar transform W_n^{-1}, obtaining a matrix B, and then convert the *columns* of B to their Haar coefficients, using the matrix W_m^{-1}.

Because columns and rows are exchanged in the first step,

$$B = A(W_n^{-1})^\top,$$

and in the second step $C = W_m^{-1}B$, thus, we have

$$C = W_m^{-1}A(W_n^{-1})^\top = D_m W_m^\top A W_n \, D_n.$$

In the other direction, given a matrix C of Haar coefficients, we reconstruct the matrix A (the image) by first applying W_m to the columns of C, obtaining B, and then $W_n^\top$ to the rows of B. Therefore

$$A = W_m C W_n^\top.$$

Of course, we dont actually have to invert W_m and W_n and perform matrix multiplications. We just have to use our algorithms using averaging and differencing.

Here is an example. If the data matrix (the image) is the 8×8 matrix

$$A = \begin{pmatrix} 64 & 2 & 3 & 61 & 60 & 6 & 7 & 57 \\ 9 & 55 & 54 & 12 & 13 & 51 & 50 & 16 \\ 17 & 47 & 46 & 20 & 21 & 43 & 42 & 24 \\ 40 & 26 & 27 & 37 & 36 & 30 & 31 & 33 \\ 32 & 34 & 35 & 29 & 28 & 38 & 39 & 25 \\ 41 & 23 & 22 & 44 & 45 & 19 & 18 & 48 \\ 49 & 15 & 14 & 52 & 53 & 11 & 10 & 56 \\ 8 & 58 & 59 & 5 & 4 & 62 & 63 & 1 \end{pmatrix},$$

then applying our algorithms, we find that

$$C = \begin{pmatrix} 32.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & -4 & 4 & -4 \\ 0 & 0 & 0 & 0 & 4 & -4 & 4 & -4 \\ 0 & 0 & 0.5 & 0.5 & 27 & -25 & 23 & -21 \\ 0 & 0 & -0.5 & -0.5 & -11 & 9 & -7 & 5 \\ 0 & 0 & 0.5 & 0.5 & -5 & 7 & -9 & 11 \\ 0 & 0 & -0.5 & -0.5 & 21 & -23 & 25 & -27 \end{pmatrix}.$$

As we can see, C has a more zero entries than A; it is a compressed version of A. We can further compress C by setting to 0 all entries of absolute value at most 0.5. Then, we get

$$C_2 = \begin{pmatrix} 32.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & -4 & 4 & -4 \\ 0 & 0 & 0 & 0 & 4 & -4 & 4 & -4 \\ 0 & 0 & 0 & 0 & 27 & -25 & 23 & -21 \\ 0 & 0 & 0 & 0 & -11 & 9 & -7 & 5 \\ 0 & 0 & 0 & 0 & -5 & 7 & -9 & 11 \\ 0 & 0 & 0 & 0 & 21 & -23 & 25 & -27 \end{pmatrix}.$$

We find that the reconstructed image is

$$A_2 = \begin{pmatrix} 63.5 & 1.5 & 3.5 & 61.5 & 59.5 & 5.5 & 7.5 & 57.5 \\ 9.5 & 55.5 & 53.5 & 11.5 & 13.5 & 51.5 & 49.5 & 15.5 \\ 17.5 & 47.5 & 45.5 & 19.5 & 21.5 & 43.5 & 41.5 & 23.5 \\ 39.5 & 25.5 & 27.5 & 37.5 & 35.5 & 29.5 & 31.5 & 33.5 \\ 31.5 & 33.5 & 35.5 & 29.5 & 27.5 & 37.5 & 39.5 & 25.5 \\ 41.5 & 23.5 & 21.5 & 43.5 & 45.5 & 19.5 & 17.5 & 47.5 \\ 49.5 & 15.5 & 13.5 & 51.5 & 53.5 & 11.5 & 9.5 & 55.5 \\ 7.5 & 57.5 & 59.5 & 5.5 & 3.5 & 61.5 & 63.5 & 1.5 \end{pmatrix},$$

which is pretty close to the original image matrix A.

It turns out that **`Matlab`** has a wonderful command, `image(X)`, which displays the matrix X has an image.

The images corresponding to A and C are shown in Figure 1.15. The compressed images corresponding to A_2 and C_2 are shown in Figure 1.16.

The compressed versions appear to be indistinguishable from the originals!

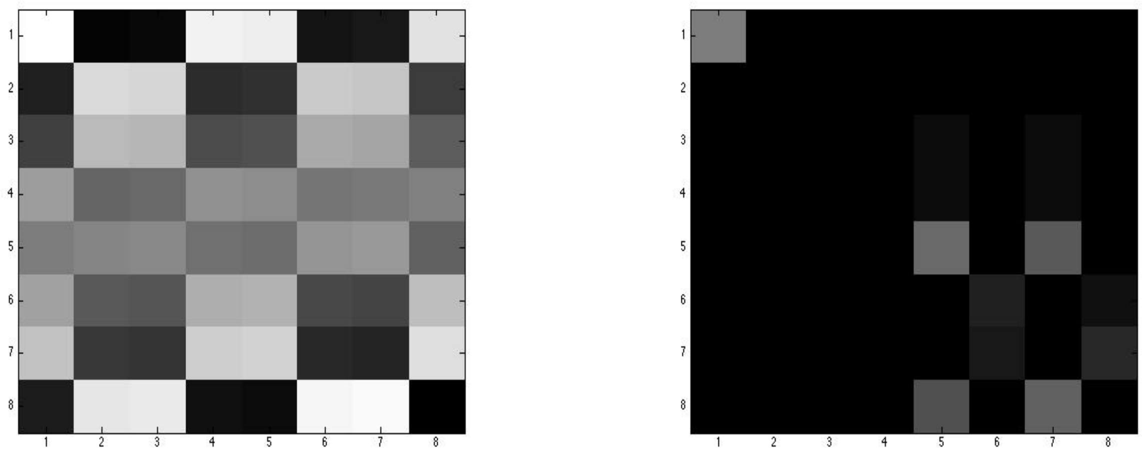


Figure 1.15: An image and its Haar transform

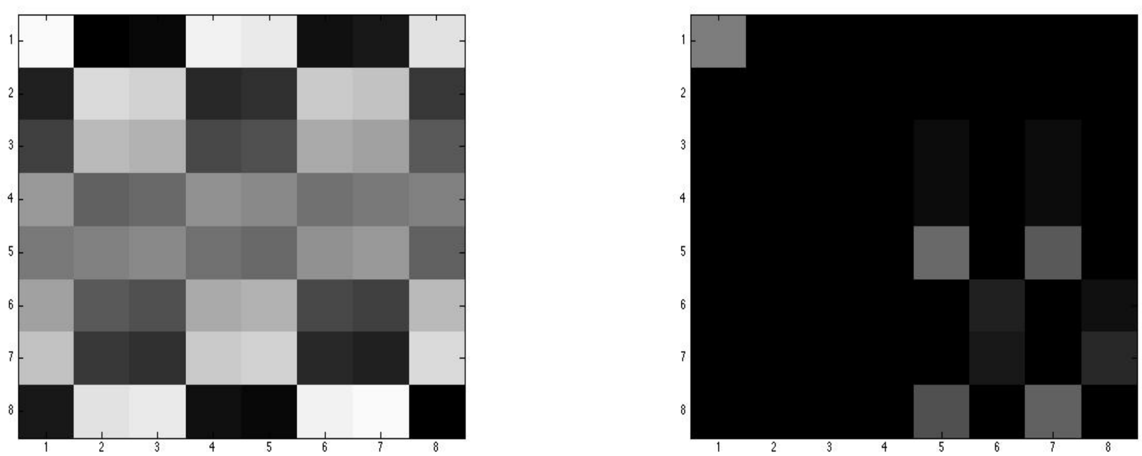


Figure 1.16: Compressed image and its Haar transform

If we use the normalized matrices H_m and H_n, then the equations relating the image matrix A and its normalized Haar transform C are

$$\begin{aligned} C &= H_m^\top A H_n \\ A &= H_m C H_n^\top. \end{aligned}$$

The Haar transform can also be used to send large images progressively over the internet.

Observe that instead of performing all rounds of averaging and differencing on each row and each column, we can perform partial encoding (and decoding).

For example, we can perform a single round of averaging and differencing for each row and each column.

The result is an image consisting of four subimages, where the top left quarter is a coarser version of the original, and the rest (consisting of three pieces) contain the finest detail coefficients.

We can also perform two rounds of averaging and differencing, or three rounds, *etc.* This process is illustrated on the image shown in Figure 1.17. The result of performing

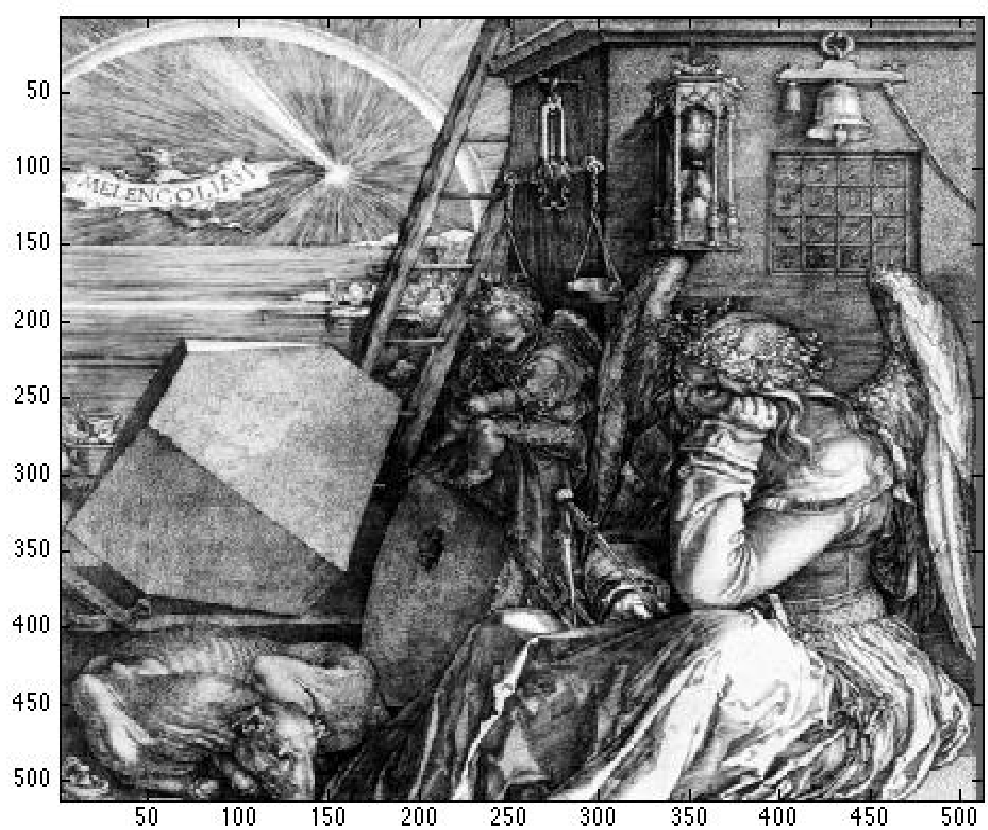


Figure 1.17: Original drawing by Durer

one round, two rounds, three rounds, and nine rounds of averaging is shown in Figure 1.18.

Since our images have size 512×512, nine rounds of averaging yields the Haar transform, displayed as the image on the bottom right. The original image has completely disappeared!

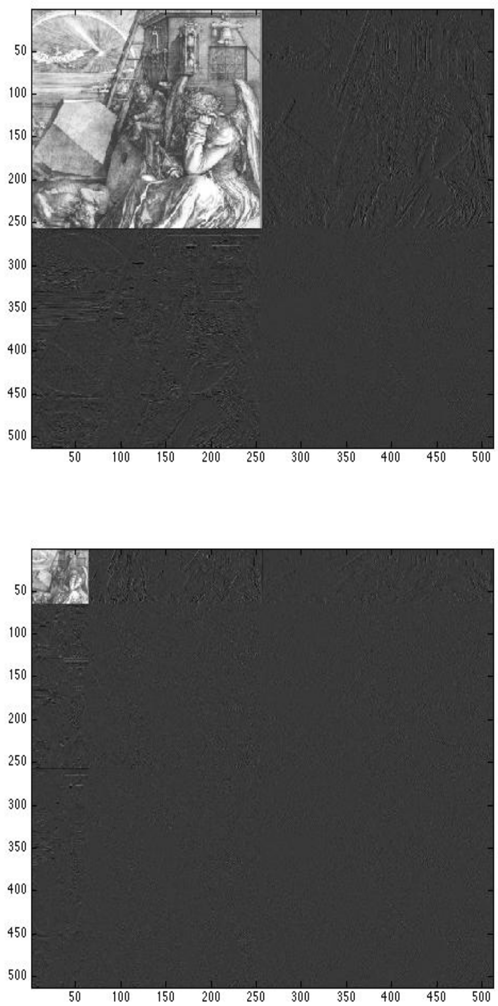


Figure 1.18: Haar tranforms after one, two, three, and nine rounds of averaging

We can find easily a basis of $2^n \times 2^n = 2^{2n}$ vectors w_{ij} for the linear map that reconstructs an image from its Haar coefficients, in the sense that for any matrix C of Haar coefficients, the image matrix A is given by

$$A = \sum_{i=1}^{2^n} \sum_{j=1}^{2^n} c_{ij} w_{ij}.$$

Indeed, the matrix w_j is given by the so-called outer product

$$w_{ij} = w_i (w_j)^\top.$$

Similarly, there is a basis of $2^n \times 2^n = 2^{2n}$ vectors h_{ij} for the 2D Haar transform, in the sense that for any matrix A, its matrix C of Haar coefficients is given by

$$C = \sum_{i=1}^{2^n} \sum_{j=1}^{2^n} a_{ij} h_{ij}.$$

If $W^{-1} = (w_{ij}^{-1})$, then

$$h_{ij} = w_i^{-1} (w_j^{-1})^\top.$$

1.8 The Effect of a Change of Bases on Matrices

The effect of a change of bases on the representation of a linear map is described in the following proposition.

Proposition 1.15. *Let E and F be vector spaces, let $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{U}' = (u'_1, \ldots, u'_n)$ be two bases of E, and let $\mathcal{V} = (v_1, \ldots, v_m)$ and $\mathcal{V}' = (v'_1, \ldots, v'_m)$ be two bases of F. Let $P = P_{\mathcal{U}',\mathcal{U}}$ be the change of basis matrix from $\mathcal{U}$ to $\mathcal{U}'$, and let $Q = P_{\mathcal{V}',\mathcal{V}}$ be the change of basis matrix from $\mathcal{V}$ to $\mathcal{V}'$. For any linear map $f\colon E \to F$, let $M(f) = M_{\mathcal{U},\mathcal{V}}(f)$ be the matrix associated to f w.r.t. the bases $\mathcal{U}$ and $\mathcal{V}$, and let $M'(f) = M_{\mathcal{U}',\mathcal{V}'}(f)$ be the matrix associated to f w.r.t. the bases $\mathcal{U}'$ and $\mathcal{V}'$. We have*

$$M'(f) = Q^{-1}M(f)P,$$

or more explicitly

$$M_{\mathcal{U}',\mathcal{V}'}(f) = P^{-1}_{\mathcal{V}',\mathcal{V}}M_{\mathcal{U},\mathcal{V}}(f)P_{\mathcal{U}',\mathcal{U}} = P_{\mathcal{V},\mathcal{V}'}M_{\mathcal{U},\mathcal{V}}(f)P_{\mathcal{U}',\mathcal{U}}.$$

As a corollary, we get the following result.

Corollary 1.16. *Let E be a vector space, and let $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{U}' = (u'_1, \ldots, u'_n)$ be two bases of E. Let $P = P_{\mathcal{U}',\mathcal{U}}$ be the change of basis matrix from $\mathcal{U}$ to $\mathcal{U}'$. For any linear map $f \colon E \to E$, let $M(f) = M_{\mathcal{U}}(f)$ be the matrix associated to f w.r.t. the basis $\mathcal{U}$, and let $M'(f) = M_{\mathcal{U}'}(f)$ be the matrix associated to f w.r.t. the basis $\mathcal{U}'$. We have*

$$M'(f) = P^{-1}M(f)P,$$

or more explicitly,

$$M_{\mathcal{U}'}(f) = P_{\mathcal{U}',\mathcal{U}}^{-1} M_{\mathcal{U}}(f) P_{\mathcal{U}',\mathcal{U}} = P_{\mathcal{U},\mathcal{U}'} M_{\mathcal{U}}(f) P_{\mathcal{U}',\mathcal{U}}.$$

Example 1.11. Let $E = \mathbb{R}^2$, $\mathcal{U} = (e_1, e_2)$ where $e_1 = (1, 0)$ and $e_2 = (0, 1)$ are the canonical basis vectors, let $\mathcal{V} = (v_1, v_2) = (e_1, e_1 - e_2)$, and let

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

The change of basis matrix $P = P_{\mathcal{V},\mathcal{U}}$ from $\mathcal{U}$ to $\mathcal{V}$ is

$$P = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix},$$

and we check that $P^{-1} = P$.

Therefore, in the basis $\mathcal{V}$, the matrix representing the linear map f defined by A is

$$\begin{aligned} A' = P^{-1}AP = PAP &= \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = D, \end{aligned}$$

a diagonal matrix.

Therefore, in the basis $\mathcal{V}$, it is clear what the action of f is: it is a stretch by a factor of 2 in the v_1 direction and it is the identity in the v_2 direction.

Observe that v_1 and v_2 are not orthogonal.

What happened is that we *diagonalized* the matrix A.

The diagonal entries 2 and 1 are the *eigenvalues* of A (and f) and v_1 and v_2 are corresponding *eigenvectors*.

The above example showed that the same linear map can be represented by different matrices. This suggests making the following definition:

Definition 1.12. Two $n \times n$ matrices A and B are said to be *similar* iff there is some invertible matrix P such that

$$B = P^{-1}AP.$$

It is easily checked that similarity is an equivalence relation.

From our previous considerations, two $n \times n$ matrices A and B are similar iff they represent the same linear map with respect to two different bases.

The following surprising fact can be shown: *Every square matrix A is similar to its transpose $A^{\top}$.*

The proof requires advanced concepts than we will not discuss in these notes (the Jordan form, or similarity invariants).

If $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{V} = (v_1, \ldots, v_n)$ are two bases of E, the change of basis matrix

$$P = P_{\mathcal{V},\mathcal{U}} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

from $(u_1, \ldots, u_n)$ to $(v_1, \ldots, v_n)$ is the matrix whose jth column consists of the coordinates of v_j over the basis $(u_1, \ldots, u_n)$, which means that

$$v_j = \sum_{i=1}^{n} a_{ij} u_i.$$

It is natural to extend the matrix notation and to express the vector $(v_1, \ldots, v_n)$ in E^n as the product of a matrix times the vector $(u_1, \ldots, u_n)$ in E^n, namely as

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix},$$

but notice that *the matrix involved is not P, but its transpose $P^\top$*.

This observation has the following consequence: if $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{V} = (v_1, \ldots, v_n)$ are two bases of E and if

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = A \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

that is,

$$v_i = \sum_{j=1}^{n} a_{ij} u_j,$$

for any vector $w \in E$, if

$$w = \sum_{i=1}^{n} x_i u_i = \sum_{k=1}^{n} y_k v_k,$$

then

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A^\top \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

and so

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = (A^\top)^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

It is easy to see that $(A^\top)^{-1} = (A^{-1})^\top$.

Also, if $\mathcal{U} = (u_1, \ldots, u_n)$, $\mathcal{V} = (v_1, \ldots, v_n)$, and $\mathcal{W} = (w_1, \ldots, w_n)$ are three bases of E, and if the change of basis matrix from $\mathcal{U}$ to $\mathcal{V}$ is $P = P_{\mathcal{V},\mathcal{U}}$ and the change of basis matrix from $\mathcal{V}$ to $\mathcal{W}$ is $Q = P_{\mathcal{W},\mathcal{V}}$, then

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = P^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \quad \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = Q^\top \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix},$$

so

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = Q^\top P^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = (PQ)^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

which means that the change of basis matrix $P_{\mathcal{W},\mathcal{U}}$ from $\mathcal{U}$ to $\mathcal{W}$ is PQ.

This proves that

$$P_{\mathcal{W},\mathcal{U}} = P_{\mathcal{V},\mathcal{U}} P_{\mathcal{W},\mathcal{V}}.$$

Even though matrices are indispensable since they are *the major tool* in applications of linear algebra, one should not lose track of the fact that

linear maps are more fundamental, because they are intrinsic objects that do not depend on the choice of bases. Consequently, we advise the reader to try to think in terms of linear maps rather than reduce everthing to matrices.

In our experience, this is particularly effective when it comes to proving results about linear maps and matrices, where proofs involving linear maps are often more "conceptual."

Also, instead of thinking of a matrix decomposition, as a purely algebraic operation, it is often illuminating to view it as a *geometric decomposition*.

After all, a

a matrix is a representation of a linear map

and most decompositions of a matrix reflect the fact that with a *suitable choice of a basis (or bases)*, the linear map is a represented by a matrix having a special shape.

The problem is then to find such bases.

Also, always try to keep in mind that

linear maps are geometric in nature; they act on space.

1.9 Affine Maps

We showed in Section 1.5 that every linear map f must send the zero vector to the zero vector, that is,

$$f(0) = 0.$$

Yet, for any fixed nonzero vector $u \in E$ (where E is any vector space), the function t_u given by

$$t_u(x) = x + u, \quad \text{for all} \quad x \in E$$

shows up in pratice (for example, in robotics).

Functions of this type are called *translations*. They are *not* linear for $u \neq 0$, since $t_u(0) = 0 + u = u$.

More generally, functions combining linear maps and translations occur naturally in many applications (robotics, computer vision, etc.), so it is necessary to understand some basic properties of these functions.

For this, the notion of affine combination turns out to play a key role.

Recall from Section 1.5 that for any vector space E, given any family $(u_i)_{i \in I}$ of vectors $u_i \in E$, an *affine combination* of the family $(u_i)_{i \in I}$ is an expression of the form

$$\sum_{i \in I} \lambda_i u_i \quad \text{with} \quad \sum_{i \in I} \lambda_i = 1,$$

where $(\lambda_i)_{i \in I}$ is a family of scalars.

A linear combination is always an affine combination, but an affine combination is a linear combination, *with the restriction that the scalars λ_i must add up to 1*.

Affine combinations are also called *barycentric combinations*.

Although this is not obvious at first glance, the condition that the scalars λ_i add up to 1 ensures that affine combinations are preserved under translations.

To make this precise, consider functions $f\colon E \to F$, where E and F are two vector spaces, such that there is some *linear map* $h\colon E \to F$ and some fixed vector $b \in F$ (a *translation vector*), such that

$$f(x) = h(x) + b, \quad \text{for all} \quad x \in E.$$

The map f given by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 8/5 & -6/5 \\ 3/10 & 2/5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

is an example of the composition of a linear map with a translation.

We claim that functions of this type preserve affine combinations.

Proposition 1.17. *For any two vector spaces E and F, given any function $f\colon E \to F$ defined such that*

$$f(x) = h(x) + b, \quad \text{for all} \quad x \in E,$$

where $h\colon E \to F$ is a linear map and b is some fixed vector in F, for every affine combination $\sum_{i \in I} \lambda_i u_i$ (with $\sum_{i \in I} \lambda_i = 1$), we have

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

In other words, f preserves affine combinations.

Surprisingly, the converse of Proposition 1.17 also holds.

Proposition 1.18. *For any two vector spaces E and F, let $f\colon E \to F$ be any function that preserves affine combinations, i.e., for every affine combination $\sum_{i\in I} \lambda_i u_i$ (with $\sum_{i\in I} \lambda_i = 1$), we have*

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

Then, for any $a \in E$, the function $h\colon E \to F$ given by

$$h(x) = f(a + x) - f(a)$$

is a linear map independent of a, and

$$f(a + x) = f(a) + h(x), \quad \text{for all} \quad x \in E.$$

In particular, for $a = 0$, if we let $c = f(0)$, then

$$f(x) = c + h(x), \quad \text{for all} \quad x \in E.$$

We should think of a as a *chosen origin* in E.

The function f maps the origin a in E to the origin $f(a)$ in F.

Proposition 1.18 shows that the definition of h does not depend on the origin chosen in E. Also, since

$$f(x) = c + h(x), \quad \text{for all} \quad x \in E$$

for some fixed vector $c \in F$, we see that f is the composition of the linear map h with the translation t_c (in F).

The unique linear map h as above is called the *linear map associated with f* and it is sometimes denoted by $\overrightarrow{f}$.

Observe that the linear map associated with a pure translation is the identity.

In view of Propositions 1.17 and 1.18, it is natural to make the following definition.

Definition 1.13. For any two vector spaces E and F, a function $f\colon E \to F$ is an *affine map* if f preserves affine combinations, *i.e.*, for every affine combination $\sum_{i\in I} \lambda_i u_i$ (with $\sum_{i\in I} \lambda_i = 1$), we have

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

Equivalently, a function $f\colon E \to F$ is an *affine map* if there is some linear map $h\colon E \to F$ (also denoted by $\overrightarrow{f}$) and some fixed vector $c \in F$ such that

$$f(x) = c + h(x), \quad \text{for all} \quad x \in E.$$

Note that a linear map always maps the standard origin 0 in E to the standard origin 0 in F.

However an affine map usually maps 0 to a nonzero vector $c = f(0)$. This is the "translation component" of the affine map.

When we deal with affine maps, it is often fruitful to think of the elements of E and F not only as vectors but also as *points*.

In this point of view, *points can only be combined using affine combinations*, but vectors can be combined in an unrestricted fashion using linear combinations.

We can also think of $u + v$ as the *result of translating the point u by the translation t_v*.

These ideas lead to the definition of *affine spaces*, but this would lead us to far afield, and for our purposes, it is enough to stick to vector spaces.

Still, one should be aware that affine combinations really apply to points, and that points are not vectors!

If E and F are finite dimensional vector spaces, with $\dim(E) = n$ and $\dim(F) = m$, then it is useful to represent an affine map with respect to bases in E in F.

However, the translation part c of the affine map must be somewhow incorporated.

There is a standard trick to do this which amounts to viewing an affine map as a linear map between spaces of dimension $n+1$ and $m+1$.

We also have the extra flexibility of choosing origins, $a \in E$ and $b \in F$.

Let $(u_1, \ldots, u_n)$ be a basis of E, $(v_1, \ldots, v_m)$ be a basis of F, and let $a \in E$ and $b \in F$ be any two fixed vectors viewed as *origins*.

Our affine map f has the property that

$$f(a+x) = c + h(x), \quad x \in E.$$

Thus, using our origins a and b, we can write

$$f(a+x) - b = c - b + h(x), \quad x \in E.$$

Over the basis $(u_1, \ldots, u_n)$, we write

$$x = x_1 u_1 + \cdots + x_n u_n,$$

and over the basis $(v_1, \ldots, v_m)$, we write

$$y = y_1 v_1 + \cdots + y_m v_m.$$

We also write

$$d = c - b = d_1 v_1 + \cdots + d_m v_m.$$

Then, with $y = f(a + x) - b$, we have

$$y = h(x) + d.$$

If we let A be the $m \times n$ matrix representing the linear map h, that is, the jth column of A consists of the coordinates of $h(u_j)$ over the basis $(v_1, \ldots, v_m)$, then we can write

$$y = Ax + d, \quad x \in \mathbb{R}^n.$$

This is the *matrix representation* of our affine map f.

The reason for using the origins a and b is that it gives us more flexibility.

In particular, when $E = F$, if there is some $a \in E$ such that $f(a) = a$ (a is a *fixed point* of f), then we can pick $b = a$.

Then, because $f(a) = a$, we get

$$v = f(u) = f(a+u-a) = f(a)+h(u-a) = a+h(u-a),$$

that is

$$v - a = h(u - a).$$

With respect to the new origin a, if we define x and y by

$$\begin{aligned} x &= u - a \\ y &= v - a, \end{aligned}$$

then we get

$$y = h(x).$$

Then, f really behaves like a linear map, but *with respect to the new origin a* (not the *standard origin* 0). This is the case of a rotation around an axis that does not pass through the origin.

Remark: A pair $(a, (u_1, \ldots, u_n))$ where $(u_1, \ldots, u_n)$ is a basis of E and a is an origin chosen in E is called an *affine frame*.

We now describe the trick which allows us to incorporate the translation part d into the matrix A.

We define the $(m+1)\times(n+1)$ matrix A' obtained by first adding d as the $(n+1)$th column, and then $(\underbrace{0, \ldots, 0}_{n}, 1)$ as the $(m+1)$th row:

$$A' = \begin{pmatrix} A & d \\ 0_n & 1 \end{pmatrix}.$$

Then, it is clear that

$$\begin{pmatrix} y \\ 1 \end{pmatrix} = \begin{pmatrix} A & d \\ 0_n & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix}$$

iff

$$y = Ax + d.$$

This amounts to considering a point $x \in \mathbb{R}^n$ as a point $(x, 1)$ *in the (affine) hyperplane H_{n+1} in $\mathbb{R}^{n+1}$ of equation $x_{n+1} = 1$*.

Then, an affine map is the restriction to the hyperplane H_{n+1} of the linear map $\widehat{f}$ from $\mathbb{R}^{n+1}$ to $\mathbb{R}^{m+1}$ corresponding to the matrix A', which maps H_{n+1} into H_{m+1} ($\widehat{f}(H_{n+1}) \subseteq H_{m+1}$).

Figure 1.19 illustrates this process for $n = 2$.

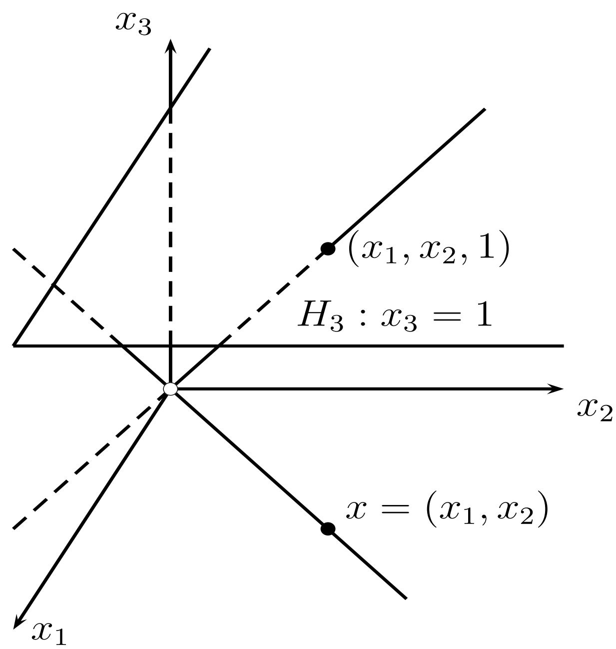


Figure 1.19: Viewing $\mathbb{R}^n$ as a hyperplane in $\mathbb{R}^{n+1}$ $(n = 2)$

For example, the map

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

defines an affine map f which is represented in $\mathbb{R}^3$ by

$$\begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 & 3 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}.$$

It is easy to check that the point $a = (6, -3)$ is fixed by f, which means that $f(a) = a$, so by translating the coordinate frame to the origin a, the affine map behaves like a linear map.

The idea of considering $\mathbb{R}^n$ as an hyperplane in $\mathbb{R}^{n+1}$ can be used to define *projective maps*.

Figure 1.20: Dog Logic

1.10 Direct Products, Sums, and Direct Sums

There are some useful ways of forming new vector spaces from older ones.

Definition 1.14. Given $p \geq 2$ vector spaces $E_1, \ldots, E_p$, the product $F = E_1 \times \cdots \times E_p$ can be made into a vector space by defining addition and scalar multiplication as follows:

$$\begin{aligned}(u_1, \ldots, u_p) + (v_1, \ldots, v_p) &= (u_1 + v_1, \ldots, u_p + v_p) \\ \lambda(u_1, \ldots, u_p) &= (\lambda u_1, \ldots, \lambda u_p),\end{aligned}$$

for all $u_i, v_i \in E_i$ and all $\lambda \in \mathbb{R}$.

With the above addition and multiplication, the vector space $F = E_1 \times \cdots \times E_p$ is called the *direct product* of the vector spaces $E_1, \ldots, E_p$.

The *projection maps* $pr_i \colon E_1 \times \cdots \times E_p \to E_i$ given by

$$pr_i(u_1, \ldots, u_p) = u_i$$

are clearly linear.

Similarly, the maps $\mathrm{in}_i \colon E_i \to E_1 \times \cdots \times E_p$ given by

$$\mathrm{in}_i(u_i) = (0, \ldots, 0, u_i, 0, \ldots, 0)$$

are injective and linear.

It can be shown (using bases) that

$$\dim(E_1 \times \cdots \times E_p) = \dim(E_1) + \cdots + \dim(E_p).$$

Let us now consider a vector space E and p subspaces $U_1, \ldots, U_p$ of E.

We have a map

$$a \colon U_1 \times \cdots \times U_p \to E$$

given by

$$a(u_1, \ldots, u_p) = u_1 + \cdots + u_p,$$

with $u_i \in U_i$ for $i = 1, \ldots, p$.

It is clear that this map is linear, and so its image is a subspace of E denoted by

$$U_1 + \cdots + U_p$$

and called the *sum* of the subspaces $U_1, \ldots, U_p$.

By definition,

$$U_1 + \cdots + U_p = \{u_1 + \cdots + u_p \mid u_i \in U_i,\ 1 \leq i \leq p\},$$

and it is immediately verified that $U_1 + \cdots + U_p$ is the smallest subspace of E containing $U_1, \ldots, U_p$.

If the map a is injective, then $\mathrm{Ker}\, a = 0$, which means that if $u_i \in U_i$ for $i = 1, \ldots, p$ and if

$$u_1 + \cdots + u_p = 0$$

then $u_1 = \cdots = u_p = 0$.

In this case, every $u \in U_1 + \cdots + U_p$ has a *unique* expression as a sum

$$u = u_1 + \cdots + u_p,$$

with $u_i \in U_i$, for $i = 1, \ldots, p$.

It is also clear that for any p nonzero vectors $u_i \in U_i$, $u_1, \ldots, u_p$ are linearly independent.

Definition 1.15. For any vector space E and any $p \geq 2$ subspaces $U_1, \ldots, U_p$ of E, if the map a defined above is injective, then the sum $U_1 + \cdots + U_p$ is called a *direct sum* and it is denoted by

$$U_1 \oplus \cdots \oplus U_p.$$

The space E is the *direct sum* of the subspaces U_i if

$$E = U_1 \oplus \cdots \oplus U_p.$$

Observe that when the map a is injective, then it is a linear isomorphism between $U_1 \times \cdots \times U_p$ and $U_1 \oplus \cdots \oplus U_p$.

The difference is that $U_1 \times \cdots \times U_p$ is defined even if the spaces U_i are not assumed to be subspaces of some common space.

There are natural injections from each U_i to E denoted by $\mathrm{in}_i \colon U_i \to E$.

Now, if $p = 2$, it is easy to determine the kernel of the map $a\colon U_1 \times U_2 \to E$. We have

$$a(u_1, u_2) = u_1 + u_2 = 0 \text{ iff } u_1 = -u_2,\ u_1 \in U_1, u_2 \in U_2,$$

which implies that

$$\operatorname{Ker} a = \{(u, -u) \mid u \in U_1 \cap U_2\}.$$

Now, $U_1 \cap U_2$ is a subspace of E and the linear map $u \mapsto (u, -u)$ is clearly an isomorphism, so $\operatorname{Ker} a$ is isomorphic to $U_1 \cap U_2$.

As a consequence, we get the following result:

Proposition 1.19. *Given any vector space E and any two subspaces U_1 and U_2, the sum $U_1 + U_2$ is a direct sum iff $U_1 \cap U_2 = 0$.*

Because of the isomorphism

$$U_1 \times \cdots \times U_p \approx U_1 \oplus \cdots \oplus U_p,$$

we have

$$\dim(U_1 \oplus \cdots \oplus U_p) = \dim(U_1) + \cdots + \dim(U_p).$$

If E is a direct sum

$$E = U_1 \oplus \cdots \oplus U_p,$$

since every $u \in E$ can be written in a unique way as

$$u = u_1 + \cdots + u_p$$

for some $u_i \in U_i$ for $i = 1 \ldots, p$, we can define the maps $\pi_i \colon E \to U_i$, called *projections*, by

$$\pi_i(u) = \pi_i(u_1 + \cdots + u_p) = u_i.$$

It is easy to check that these maps are linear and satisfy the following properties:

$$\pi_j \circ \pi_i = \begin{cases} \pi_i & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$
$$\pi_1 + \cdots + \pi_p = \mathrm{id}_E.$$

A function f such that $f \circ f = f$ is said to be *idempotent*. Thus, the projections π_i are idempotent.

Conversely, the following proposition can be shown:

Proposition 1.20. *Let E be a vector space. For any $p \geq 2$ linear maps $f_i \colon E \to E$, if*

$$f_j \circ f_i = \begin{cases} f_i & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$
$$f_1 + \cdots + f_p = \mathrm{id}_E,$$

then if we let $U_i = f_i(E)$, we have a direct sum

$$E = U_1 \oplus \cdots \oplus U_p.$$

We also have the following proposition characterizing idempotent linear maps whose proof is also left as an exercise.

Proposition 1.21. *For every vector space E, if $f \colon E \to E$ is an idempotent linear map, i.e., $f \circ f = f$, then we have a direct sum*

$$E = \mathrm{Ker}\, f \oplus \mathrm{Im}\, f,$$

so that f is the projection onto its image $\mathrm{Im}\, f$.

We are now ready to prove a very crucial result relating the rank and the dimension of the kernel of a linear map.

Theorem 1.22. *Let $f\colon E \to F$ be a linear map. For any choice of a basis $(f_1, \ldots, f_r)$ of $\mathrm{Im}\, f$, let $(u_1, \ldots, u_r)$ be any vectors in E such that $f_i = f(u_i)$, for $i = 1, \ldots, r$. If $s\colon \mathrm{Im}\, f \to E$ is the unique linear map defined by $s(f_i) = u_i$, for $i = 1, \ldots, r$, then s is injective, $f \circ s = \mathrm{id}$, and we have a direct sum*

$$E = \mathrm{Ker}\, f \oplus \mathrm{Im}\, s$$

as illustrated by the following diagram:

$$\mathrm{Ker}\, f \longrightarrow E = \mathrm{Ker}\, f \oplus \mathrm{Im}\, s \underset{s}{\overset{f}{\rightleftarrows}} \mathrm{Im}\, f \subseteq F.$$

As a consequence,

$$\dim(E) = \dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f) = \dim(\mathrm{Ker}\, f) + \mathrm{rk}(f).$$

Remark: The dimension $\dim(\mathrm{Ker}\, f)$ of the kernel of a linear map f is often called the *nullity* of f.

We now derive some important results using Theorem 1.22.

Proposition 1.23. *Given a vector space E, if U and V are any two subspaces of E, then*

$$\dim(U) + \dim(V) = \dim(U + V) + \dim(U \cap V),$$

an equation known as Grassmann's relation.

The Grassmann relation can be very useful to figure out whether two subspace have a nontrivial intersection in spaces of dimension > 3.

For example, it is easy to see that in $\mathbb{R}^5$, there are subspaces U and V with $\dim(U) = 3$ and $\dim(V) = 2$ such that $U \cap V = 0$

However, we can show that if $\dim(U) = 3$ and $\dim(V) = 3$, then $\dim(U \cap V) \geq 1$.

As another consequence of Proposition 1.23, if U and V are two hyperplanes in a vector space of dimension n, so that $\dim(U) = n - 1$ and $\dim(V) = n - 1$, we have

$$\dim(U \cap V) \geq n - 2,$$

and so, if $U \neq V$, then

$$\dim(U \cap V) = n - 2.$$

Proposition 1.24. *If $U_1, \ldots, U_p$ are any subspaces of a finite dimensional vector space E, then*

$$\dim(U_1 + \cdots + U_p) \leq \dim(U_1) + \cdots + \dim(U_p),$$

and

$$\dim(U_1 + \cdots + U_p) = \dim(U_1) + \cdots + \dim(U_p)$$

iff the U_is form a direct sum $U_1 \oplus \cdots \oplus U_p$.

Another important corollary of Theorem 1.22 is the following result:

Proposition 1.25. *Let E and F be two vector spaces with the same finite dimension $\dim(E) = \dim(F) = n$. For every linear map $f \colon E \to F$, the following properties are equivalent:*

(a) f is bijective.

(b) f is surjective.

(c) f is injective.

(d) $\mathrm{Ker}\, f = 0$.

One should be warned that Proposition 1.25 fails in infinite dimension.

We also have the following basic proposition about injective or surjective linear maps.

Proposition 1.26. *Let E and F be vector spaces, and let $f\colon E \to F$ be a linear map. If $f\colon E \to F$ is injective, then there is a surjective linear map $r\colon F \to E$ called a retraction, such that $r \circ f = \mathrm{id}_E$. If $f\colon E \to F$ is surjective, then there is an injective linear map $s\colon F \to E$ called a section, such that $f \circ s = \mathrm{id}_F$.*

The notion of rank of a linear map or of a matrix important, both theoretically and practically, since it is the key to the solvability of linear equations.

Proposition 1.27. *Given a linear map $f\colon E \to F$, the following properties hold:*

(i) $\mathrm{rk}(f) + \dim(\mathrm{Ker}\, f) = \dim(E)$.

(ii) $\mathrm{rk}(f) \leq \min(\dim(E), \dim(F))$.

The rank of a matrix is defined as follows.

Definition 1.16. Given a $m \times n$-matrix $A = (a_{i\,j})$, the *rank* $\mathrm{rk}(A)$ of the matrix A is the maximum number of linearly independent columns of A (viewed as vectors in $\mathbb{R}^m$).

In view of Proposition 1.4, the rank of a matrix A is the dimension of the subspace of $\mathbb{R}^m$ generated by the columns of A.

Let E and F be two vector spaces, and let $(u_1, \ldots, u_n)$ be a basis of E, and $(v_1, \ldots, v_m)$ a basis of F. Let $f \colon E \to F$ be a linear map, and let $M(f)$ be its matrix w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$.

Since the rank $\mathrm{rk}(f)$ of f is the dimension of $\mathrm{Im}\, f$, which is generated by $(f(u_1), \ldots, f(u_n))$, the rank of f is the maximum number of linearly independent vectors in $(f(u_1), \ldots, f(u_n))$, which is equal to the number of linearly independent columns of $M(f)$, since F and $\mathbb{R}^m$ are isomorphic.

Thus, we have $\mathrm{rk}(f) = \mathrm{rk}(M(f))$, for every matrix representing f.

We will see later, using duality, that the rank of a matrix A is also equal to the maximal number of linearly independent rows of A.

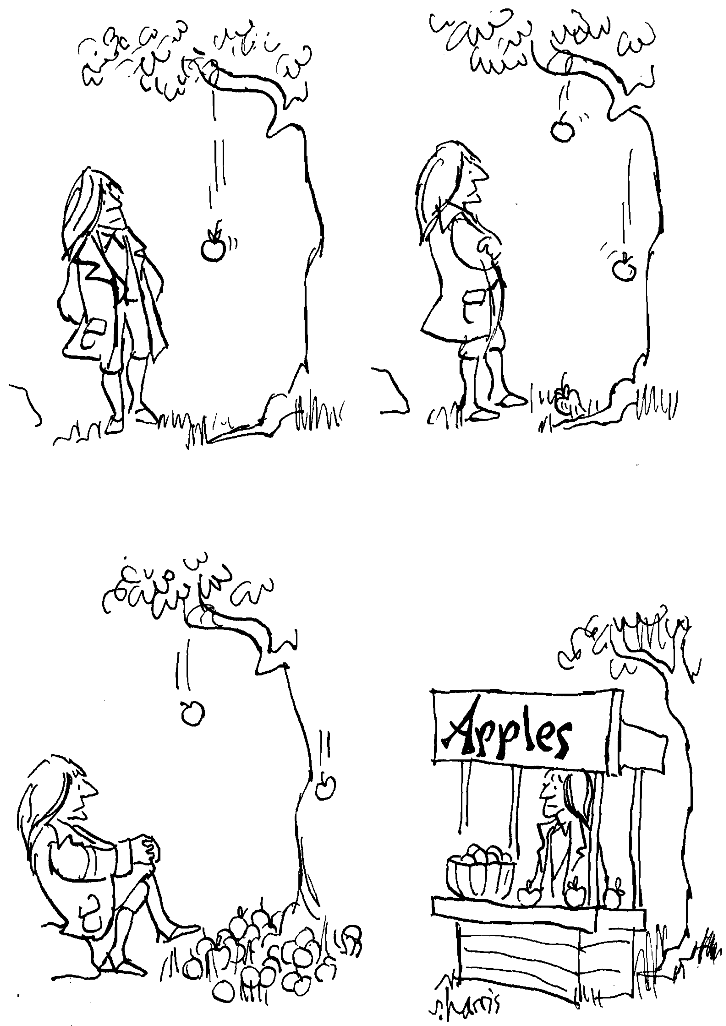


Figure 1.21: How did Newton start a business

1.11 The Dual Space E^* and Linear Forms

We already observed that the field K itself ($K = \mathbb{R}$ or $K = \mathbb{C}$) is a vector space (over itself).

The vector space $\mathrm{Hom}(E, K)$ of linear maps from E to the field K, the *linear forms*, plays a particular role.

We take a quick look at the connection between E and $E^* = \mathrm{Hom}(E, K)$, its *dual space*.

As we will see shortly, every linear map $f\colon E \to F$ gives rise to a linear map $f^\top\colon F^* \to E^*$, and it turns out that in a suitable basis, the matrix of $f^\top$ is the *transpose* of the matrix of f.

Thus, the notion of dual space provides a conceptual explanation of the phenomena associated with transposition.

But it does more, because it allows us to view subspaces as solutions of sets of linear equations and vice-versa.

Consider the following set of two "linear equations" in $\mathbb{R}^3$,

$$\begin{aligned} x - y + z &= 0 \\ x - y - z &= 0, \end{aligned}$$

and let us find out what is their set V of common solutions $(x, y, z) \in \mathbb{R}^3$.

By subtracting the second equation from the first, we get $2z = 0$, and by adding the two equations, we find that $2(x - y) = 0$, so the set V of solutions is given by

$$\begin{aligned} y &= x \\ z &= 0. \end{aligned}$$

This is a one dimensional subspace of $\mathbb{R}^3$. Geometrically, this is the line of equation $y = x$ in the plane $z = 0$.

Now, why did we say that the above equations are linear?

This is because, as functions of (x, y, z), both maps $f_1\colon (x, y, z) \mapsto x - y + z$ and $f_2\colon (x, y, z) \mapsto x - y - z$ are linear.

The set of all such linear functions from $\mathbb{R}^3$ to $\mathbb{R}$ is a vector space; we used this fact to form linear combinations of the "equations" f_1 and f_2.

Observe that the dimension of the subspace V is 1.

The ambient space has dimension $n = 3$ and there are two "independent" equations f_1, f_2, so it appears that the dimension $\dim(V)$ of the subspace V defined by m independent equations is

$$\dim(V) = n - m,$$

which is indeed a general fact.

More generally, in $\mathbb{R}^n$, a linear equation is determined by an n-tuple $(a_1, \ldots, a_n) \in \mathbb{R}^n$, and the solutions of this linear equation are given by the n-tuples $(x_1, \ldots, x_n) \in \mathbb{R}^n$ such that

$$a_1x_1 + \cdots + a_nx_n = 0;$$

these solutions constitute the kernel of the linear map $(x_1, \ldots, x_n) \mapsto a_1x_1 + \cdots + a_nx_n$.

The above considerations assume that we are working in the canonical basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$, but we can define "linear equations" independently of bases and in any dimension, by viewing them as elements of the vector space $\mathrm{Hom}(E, K)$ of linear maps from E to the field K.

Definition 1.17. Given a vector space E, the vector space $\mathrm{Hom}(E, K)$ of linear maps from E to K is called the *dual space (or dual)* of E. The space $\mathrm{Hom}(E, K)$ is also denoted by E^*, and the linear maps in E^* are called *the linear forms*, or *covectors*. The dual space E^{**} of the space E^* is called the *bidual* of E.

As a matter of notation, linear forms $f\colon E \to K$ will also be denoted by starred symbol, such as u^*, x^*, etc.

If E is a vector space of finite dimension n and $(u_1, \ldots, u_n)$ is a basis of E, for any linear form $f^* \in E^*$, for every $x = x_1u_1 + \cdots + x_nu_n \in E$, we have

$$f^*(x) = \lambda_1 x_1 + \cdots + \lambda_n x_n,$$

where $\lambda_i = f^*(u_i) \in K$, for every i, $1 \leq i \leq n$.

Thus, with respect to the basis $(u_1, \ldots, u_n)$, $f^*(x)$ is a linear combination of the coordinates of x, and we can view a linear form as a *linear equation*.

Given a linear form $u^* \in E^*$ and a vector $v \in E$, the result $u^*(v)$ of applying u^* to v is also denoted by $\langle u^*, v\rangle$.

This defines a binary operation $\langle -, - \rangle\colon E^* \times E \to K$ satisfying the following properties:

$$\begin{aligned} \langle u_1^* + u_2^*, v\rangle &= \langle u_1^*, v\rangle + \langle u_2^*, v\rangle \\ \langle u^*, v_1 + v_2\rangle &= \langle u^*, v_1\rangle + \langle u^*, v_2\rangle \\ \langle \lambda u^*, v\rangle &= \lambda\langle u^*, v\rangle \\ \langle u^*, \lambda v\rangle &= \lambda\langle u^*, v\rangle. \end{aligned}$$

The above identities mean that $\langle -, - \rangle$ is a *bilinear map*, since it is linear in each argument.

It is often called the *canonical pairing* between E^* and E.

In view of the above identities, given any fixed vector $v \in E$, the map $\mathrm{eval}_v \colon E^* \to K$ (*evaluation at v*) defined such that

$$\mathrm{eval}_v(u^*) = \langle u^*, v\rangle = u^*(v) \quad \text{for every } u^* \in E^*$$

is a linear map from E^* to K, that is, eval_v is a linear form in E^{**}.

Again from the above identities, the map $\mathrm{eval}_E \colon E \to E^{**}$, defined such that

$$\mathrm{eval}_E(v) = \mathrm{eval}_v \quad \text{for every } v \in E,$$

is a linear map.

We shall see that it is injective, and that it is an isomorphism when E has finite dimension.

We now formalize the notion of the set V^0 of linear equations vanishing on all vectors in a given subspace $V \subseteq E$, and the notion of the set U^0 of common solutions of a given set $U \subseteq E^*$ of linear equations.

The duality theorem (Theorem 1.28) shows that the dimensions of V and V^0, and the dimensions of U and U^0, are related in a crucial way.

It also shows that, in finite dimension, the maps $V \mapsto V^0$ and $U \mapsto U^0$ are inverse bijections from subspaces of E to subspaces of E^*.

Definition 1.18. Given a vector space E and its dual E^*, we say that a vector $v \in E$ and a linear form $u^* \in E^*$ are *orthogonal* iff $\langle u^*, v\rangle = 0$. Given a subspace V of E and a subspace U of E^*, we say that *V and U are orthogonal* iff $\langle u^*, v\rangle = 0$ for every $u^* \in U$ and every $v \in V$. Given a subset V of E (resp. a subset U of E^*), the *orthogonal V^0 of V* is the subspace V^0 of E^* defined such that

$$V^0 = \{u^* \in E^* \mid \langle u^*, v\rangle = 0,\ \text{for every } v \in V\}$$

(resp. the *orthogonal U^0 of U* is the subspace U^0 of E defined such that

$$U^0 = \{v \in E \mid \langle u^*, v\rangle = 0,\ \text{for every } u^* \in U\}).$$

The subspace $V^0 \subseteq E^*$ is also called the *annihilator* of V.

The subspace $U^0 \subseteq E$ annihilated by $U \subseteq E^*$ does not have a special name. It seems reasonable to call it the *linear subspace (or linear variety) defined by U*.

Informally, V^0 is the *set of linear equations that vanish on V*, and U^0 is *the set of common zeros of all linear equations in U*. We can also define V^0 by

$$V^0 = \{u^* \in E^* \mid V \subseteq \operatorname{Ker} u^*\}$$

and U^0 by

$$U^0 = \bigcap_{u^* \in U} \operatorname{Ker} u^*.$$

Observe that $E^0 = 0$, and $\{0\}^0 = E^*$.

Furthermore, if $V_1 \subseteq V_2 \subseteq E$, then $V_2^0 \subseteq V_1^0 \subseteq E^*$, and if $U_1 \subseteq U_2 \subseteq E^*$, then $U_2^0 \subseteq U_1^0 \subseteq E$.

It can also be shown that that $V \subseteq V^{00}$ for every subspace V of E, and that $U \subseteq U^{00}$ for every subspace U of E^*.

We will see shortly that in finite dimension, we have

$$V = V^{00} \quad \text{and} \quad U = U^{00}.$$

Given a vector space E and any basis $(u_i)_{i \in I}$ for E, we can associate to each u_i a linear form $u_i^* \in E^*$, and the u_i^* have some remarkable properties.

Definition 1.19. Given a vector space E and any basis $(u_i)_{i \in I}$ for E, by Proposition 1.9, for every $i \in I$, there is a unique linear form u_i^* such that

$$u_i^*(u_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

for every $j \in I$. The linear form u_i^* is called the *coordinate form* of index i w.r.t. the basis $(u_i)_{i \in I}$.

Remark: Given an index set I, authors often define the so called *Kronecker symbol* $\delta_{i\,j}$, such that

$$\delta_{i\,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

for all $i, j \in I$.

Then,

$$u_i^*(u_j) = \delta_{i\,j}.$$

The reason for the terminology *coordinate form* is as follows: If E has finite dimension and if $(u_1, \ldots, u_n)$ is a basis of E, for any vector

$$v = \lambda_1 u_1 + \cdots + \lambda_n u_n,$$

we have

$$u_i^*(v) = \lambda_i.$$

Therefore, u_i^* is the linear function that returns the ith coordinate of a vector expressed over the basis $(u_1, \ldots, u_n)$.

We have the following important duality theorem.

Theorem 1.28. *(Duality theorem) Let E be a vector space of dimension n. The following properties hold:*

(a) For every basis $(u_1, \ldots, u_n)$ of E, the family of coordinate forms $(u_1^*, \ldots, u_n^*)$ is a basis of E^*.

(b) For every subspace V of E, we have $V^{00} = V$.

(c) For every pair of subspaces V and W of E such that $E = V \oplus W$, with V of dimension m, for every basis $(u_1, \ldots, u_n)$ of E such that $(u_1, \ldots, u_m)$ is a basis of V and $(u_{m+1}, \ldots, u_n)$ is a basis of W, the family $(u_1^*, \ldots, u_m^*)$ is a basis of the orthogonal W^0 of W in E^*. Furthermore, we have $W^{00} = W$, and

$$\dim(W) + \dim(W^0) = \dim(E).$$

(d) For every subspace U of E^*, we have

$$\dim(U) + \dim(U^0) = \dim(E),$$

where U^0 is the orthogonal of U in E, and $U^{00} = U$.

Part (a) of Theorem 1.28 shows that

$$\dim(E) = \dim(E^*),$$

and if $(u_1, \ldots, u_n)$ is a basis of E, then $(u_1^*, \ldots, u_n^*)$ is a basis of the dual space E^* called the *dual basis* of $(u_1, \ldots, u_n)$.

By part (c) and (d) of theorem 1.28, the maps $V \mapsto V^0$ and $U \mapsto U^0$, where V is a subspace of E and U is a subspace of E^*, are inverse bijections.

These maps set up a *duality* between subspaces of E, and subspaces of E^*.

One should be careful that this bijection does not hold if E has infinite dimension. Some restrictions on the dimensions of U and V are needed.

When E is of finite dimension n and $(u_1, \ldots, u_n)$ is a basis of E, we noted that the family $(u_1^*, \ldots, u_n^*)$ is a basis of the dual space E^*,

Let us see how the coordinates of a linear form $\varphi^* \in E^*$ over the basis $(u_1^*, \ldots, u_n^*)$ vary under a change of basis.

Let $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ be two bases of E, and let $P = (a_{i\,j})$ be the change of basis matrix from $(u_1, \ldots, u_n)$ to $(v_1, \ldots, v_n)$, so that

$$v_j = \sum_{i=1}^{n} a_{i\,j} u_i.$$

If

$$\varphi^* = \sum_{i=1}^{n} \varphi_i u_i^* = \sum_{i=1}^{n} \varphi_i' v_i^*,$$

after some algebra, we get

$$\varphi_j' = \sum_{i=1}^{n} a_{i\,j} \varphi_i.$$

Comparing with the change of basis

$$v_j = \sum_{i=1}^{n} a_{i\,j} u_i,$$

we note that this time, the coordinates (φ_i) of the linear form φ^* change in the *same direction* as the change of basis.

For this reason, we say that the coordinates of linear forms are *covariant*.

By abuse of language, it is often said that linear forms are *covariant*, which explains why the term *covector* is also used for a linear form.

Observe that if $(e_1, \ldots, e_n)$ is a basis of the vector space E, then, as a linear map from E to K, every linear form $f \in E^*$ is represented by a $1 \times n$ matrix, that is, by a *row vector*

$$(\lambda_1 \cdots \lambda_n),$$

with respect to the basis $(e_1, \ldots, e_n)$ of E, and 1 of K, where $f(e_i) = \lambda_i$.

A vector $u = \sum_{i=1}^{n} u_i e_i \in E$ is represented by a $n \times 1$ matrix, that is, by a *column vector*

$$\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

and the action of f on u, namely $f(u)$, is represented by the matrix product

$$\begin{pmatrix} \lambda_1 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \lambda_1 u_1 + \cdots + \lambda_n u_n.$$

On the other hand, with respect to the dual basis $(e_1^*, \ldots, e_n^*)$ of E^*, the linear form f is represented by the column vector

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

We will now pin down the relationship between a vector space E and its bidual E^{**}.

Proposition 1.29. *Let E be a vector space. The following properties hold:*

(a) The linear map $\mathrm{eval}_E \colon E \to E^{**}$ defined such that

$$\mathrm{eval}_E(v) = \mathrm{eval}_v, \quad \textit{for all } v \in E,$$

that is, $\mathrm{eval}_E(v)(u^*) = \langle u^*, v\rangle = u^*(v)$ for every $u^* \in E^*$, is injective.

(b) When E is of finite dimension n, the linear map $\mathrm{eval}_E \colon E \to E^{**}$ is an isomorphism (called the canonical isomorphism).

When E is of finite dimension and $(u_1, \ldots, u_n)$ is a basis of E, in view of the canonical isomorphism $\mathrm{eval}_E \colon E \to E^{**}$, the basis $(u_1^{**}, \ldots, u_n^{**})$ of the bidual is identified with $(u_1, \ldots, u_n)$.

Proposition 1.29 can be reformulated very fruitfully in terms of pairings.

Definition 1.20. Given two vector spaces E and F over K, a *pairing between E and F* is a bilinear map $\varphi\colon E \times F \to K$. Such a pairing is *nondegenerate* iff

(1) for every $u \in E$, if $\varphi(u, v) = 0$ for all $v \in F$, then $u = 0$, and

(2) for every $v \in F$, if $\varphi(u, v) = 0$ for all $u \in E$, then $v = 0$.

A pairing $\varphi\colon E \times F \to K$ is often denoted by $\langle -, - \rangle\colon E \times F \to K$.

For example, the map $\langle -, - \rangle\colon E^* \times E \to K$ defined earlier is a nondegenerate pairing (use the proof of (a) in Proposition 1.29).

Given a pairing $\varphi\colon E \times F \to K$, we can define two maps $l_\varphi\colon E \to F^*$ and $r_\varphi\colon F \to E^*$ as follows:

For every $u \in E$, we define the linear form $l_\varphi(u)$ in F^* such that

$$l_\varphi(u)(y) = \varphi(u, y) \quad \text{for every } y \in F,$$

and for every $v \in F$, we define the linear form $r_\varphi(v)$ in E^* such that

$$r_\varphi(v)(x) = \varphi(x, v) \quad \text{for every } x \in E.$$

Proposition 1.30. *Given two vector spaces E and F over K, for every nondegenerate pairing $\varphi\colon E \times F \to K$ between E and F, the maps $l_\varphi\colon E \to F^*$ and $r_\varphi\colon F \to E^*$ are linear and injective. Furthermore, if E and F have finite dimension, then this dimension is the same and $l_\varphi\colon E \to F^*$ and $r_\varphi\colon F \to E^*$ are bijections.*

When E has finite dimension, the nondegenerate pairing $\langle -, - \rangle\colon E^* \times E \to K$ yields another proof of the existence of a natural isomorphism between E and E^{**}.

Interesting nondegenerate pairings arise in exterior algebra.

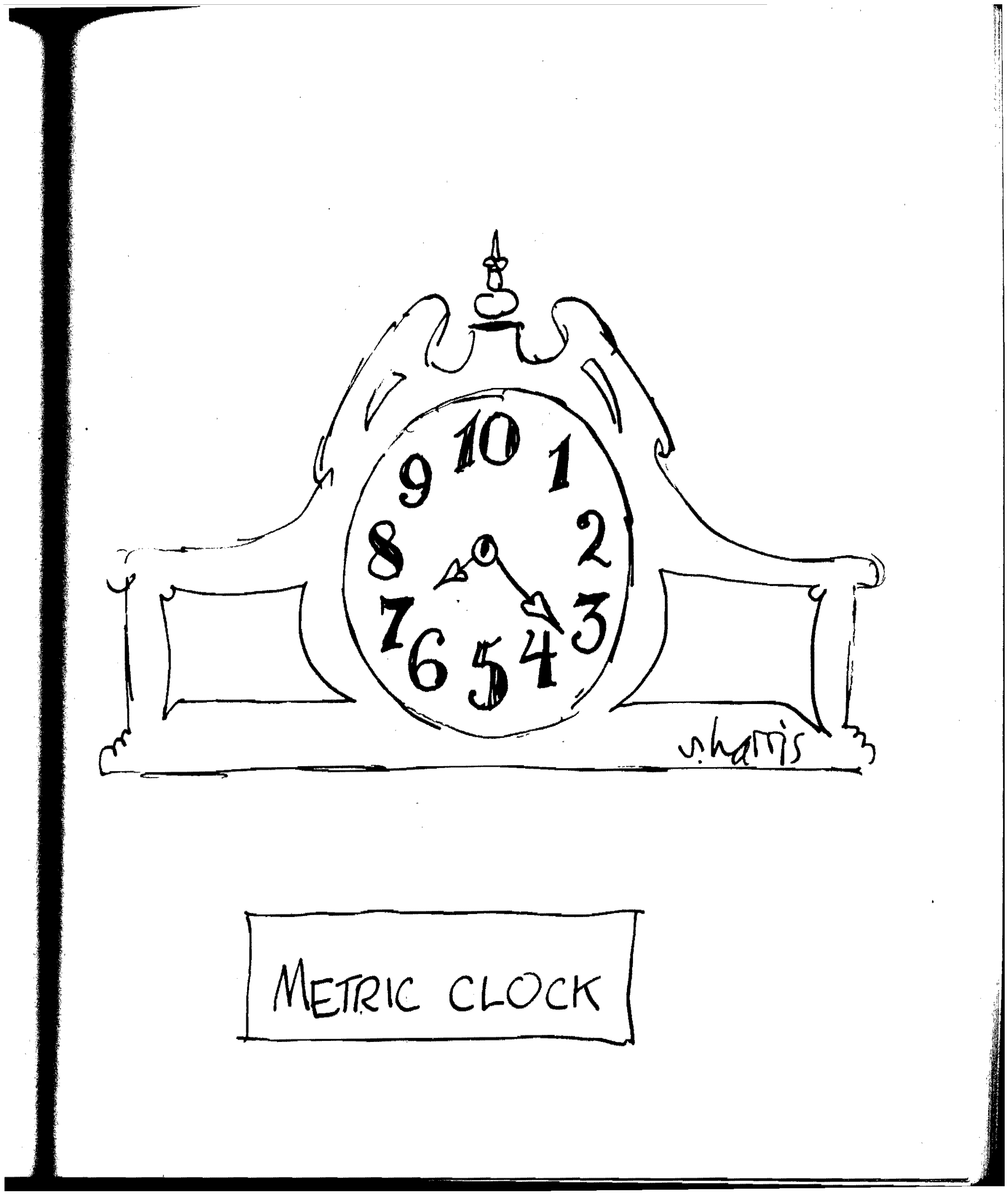


Figure 1.22: Metric Clock

1.12 Hyperplanes and Linear Forms

Actually, Proposition 1.31 below follows from parts (c) and (d) of Theorem 1.28, but we feel that it is also interesting to give a more direct proof.

Proposition 1.31. *Let E be a vector space. The following properties hold:*

(a) Given any nonnull linear form $f^* \in E^*$, its kernel $H = \mathrm{Ker}\, f^*$ is a hyperplane.

(b) For any hyperplane H in E, there is a (nonnull) linear form $f^* \in E^*$ such that $H = \mathrm{Ker}\, f^*$.

(c) Given any hyperplane H in E and any (nonnull) linear form $f^* \in E^*$ such that $H = \mathrm{Ker}\, f^*$, for every linear form $g^* \in E^*$, $H = \mathrm{Ker}\, g^*$ iff $g^* = \lambda f^*$ for some $\lambda \neq 0$ in K.

We leave as an exercise the fact that every subspace $V \neq E$ of a vector space E, is the intersection of all hyperplanes that contain V.

We now consider the notion of transpose of a linear map and of a matrix.

1.13 Transpose of a Linear Map and of a Matrix

Given a linear map $f\colon E \to F$, it is possible to define a map $f^\top\colon F^* \to E^*$ which has some interesting properties.

Definition 1.21. Given a linear map $f\colon E \to F$, the *transpose* $f^\top\colon F^* \to E^*$ of f is the linear map defined such that

$$f^\top(v^*) = v^* \circ f,$$

for every $v^* \in F^*$.

Equivalently, the linear map $f^\top\colon F^* \to E^*$ is defined such that

$$\langle v^*, f(u)\rangle = \langle f^\top(v^*), u\rangle,$$

for all $u \in E$ and all $v^* \in F^*$.

It is easy to verify that the following properties hold:

$$\begin{aligned}(f+g)^\top &= f^\top + g^\top\\ (g\circ f)^\top &= f^\top \circ g^\top\\ \mathrm{id}_E^\top &= \mathrm{id}_{E^*}.\end{aligned}$$

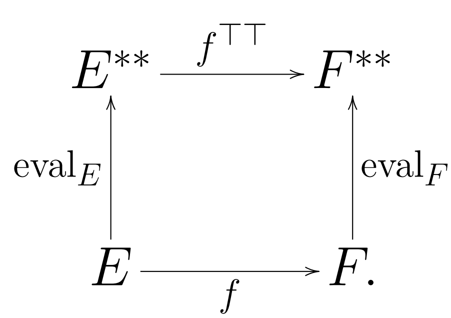
Note the reversal of composition on the right-hand side of $(g\circ f)^\top = f^\top \circ g^\top$.

We also have the following property showing the naturality of the eval map.

Proposition 1.32. *For any linear map $f\colon E\to F$, we have*

$$f^{\top\top}\circ \mathrm{eval}_E = \mathrm{eval}_F \circ f,$$

or equivalently, the following diagram commutes:



If E and F are finite-dimensional, then eval_E and eval_F are isomorphisms, and if we identify E with its bidual E^{**} and F with its bidual F^{**}, then

$$(f^\top)^\top = f.$$

Proposition 1.33. *Given a linear map $f\colon E \to F$, for any subspace V of E, we have*

$$f(V)^0 = (f^\top)^{-1}(V^0) = \{w^* \in F^* \mid f^\top(w^*) \in V^0\}.$$

As a consequence,

$$\mathrm{Ker}\, f^\top = (\mathrm{Im}\, f)^0 \quad \textit{and} \quad \mathrm{Ker}\, f = (\mathrm{Im}\, f^\top)^0.$$

The following theorem shows the relationship between the rank of f and the rank of $f^\top$.

Theorem 1.34. *Given a linear map* $f\colon E \to F$*, the following properties hold.*

(a) The dual $(\mathrm{Im}\, f)^*$ *of* $\mathrm{Im}\, f$ *is isomorphic to* $\mathrm{Im}\, f^\top = f^\top(F^*)$*; that is,*

$$(\mathrm{Im}\, f)^* \approx \mathrm{Im}\, f^\top.$$

(b) If F *is finite dimensional, then* $\mathrm{rk}(f) = \mathrm{rk}(f^\top)$*.*

The following proposition can be shown, but it requires a generalization of the duality theorem.

Proposition 1.35. *If* $f\colon E \to F$ *is any linear map, then the following identities hold:*

$$\begin{aligned}
\mathrm{Im}\, f^\top &= (\mathrm{Ker}\,(f))^0 \\
\mathrm{Ker}\,(f^\top) &= (\mathrm{Im}\, f)^0 \\
\mathrm{Im}\, f &= (\mathrm{Ker}\,(f^\top)^0 \\
\mathrm{Ker}\,(f) &= (\mathrm{Im}\, f^\top)^0.
\end{aligned}$$

The following proposition shows the relationship between the matrix representing a linear map $f\colon E \to F$ and the matrix representing its transpose $f^\top\colon F^* \to E^*$.

Proposition 1.36. *Let E and F be two vector spaces, and let $(u_1, \ldots, u_n)$ be a basis for E, and $(v_1, \ldots, v_m)$ be a basis for F. Given any linear map $f\colon E \to F$, if $M(f)$ is the $m \times n$-matrix representing f w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$, the $n \times m$-matrix $M(f^\top)$ representing $f^\top\colon F^* \to E^*$ w.r.t. the dual bases $(v_1^*, \ldots, v_m^*)$ and $(u_1^*, \ldots, u_n^*)$ is the transpose $M(f)^\top$ of $M(f)$.*

We now can give a very short proof of the fact that the rank of a matrix is equal to the rank of its transpose.

Proposition 1.37. *Given a $m \times n$ matrix A over a field K, we have $\mathrm{rk}(A) = \mathrm{rk}(A^\top)$.*

Thus, given an $m \times n$-matrix A, the maximum number of linearly independent columns is equal to the maximum number of linearly independent rows.

Proposition 1.38. *Given any $m \times n$ matrix A over a field K (typically $K = \mathbb{R}$ or $K = \mathbb{C}$), the rank of A is the maximum natural number r such that there is an invertible $r \times r$ submatrix of A obtained by selecting r rows and r columns of A.*

For example, the 3×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

has rank 2 iff one of the three 2×2 matrices

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{pmatrix} \quad \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

is invertible. We will see in Chapter 3 that this is equivalent to the fact the determinant of one of the above matrices is nonzero.

This is not a very efficient way of finding the rank of a matrix. We will see that there are better ways using various decompositions such as LU, QR, or SVD.

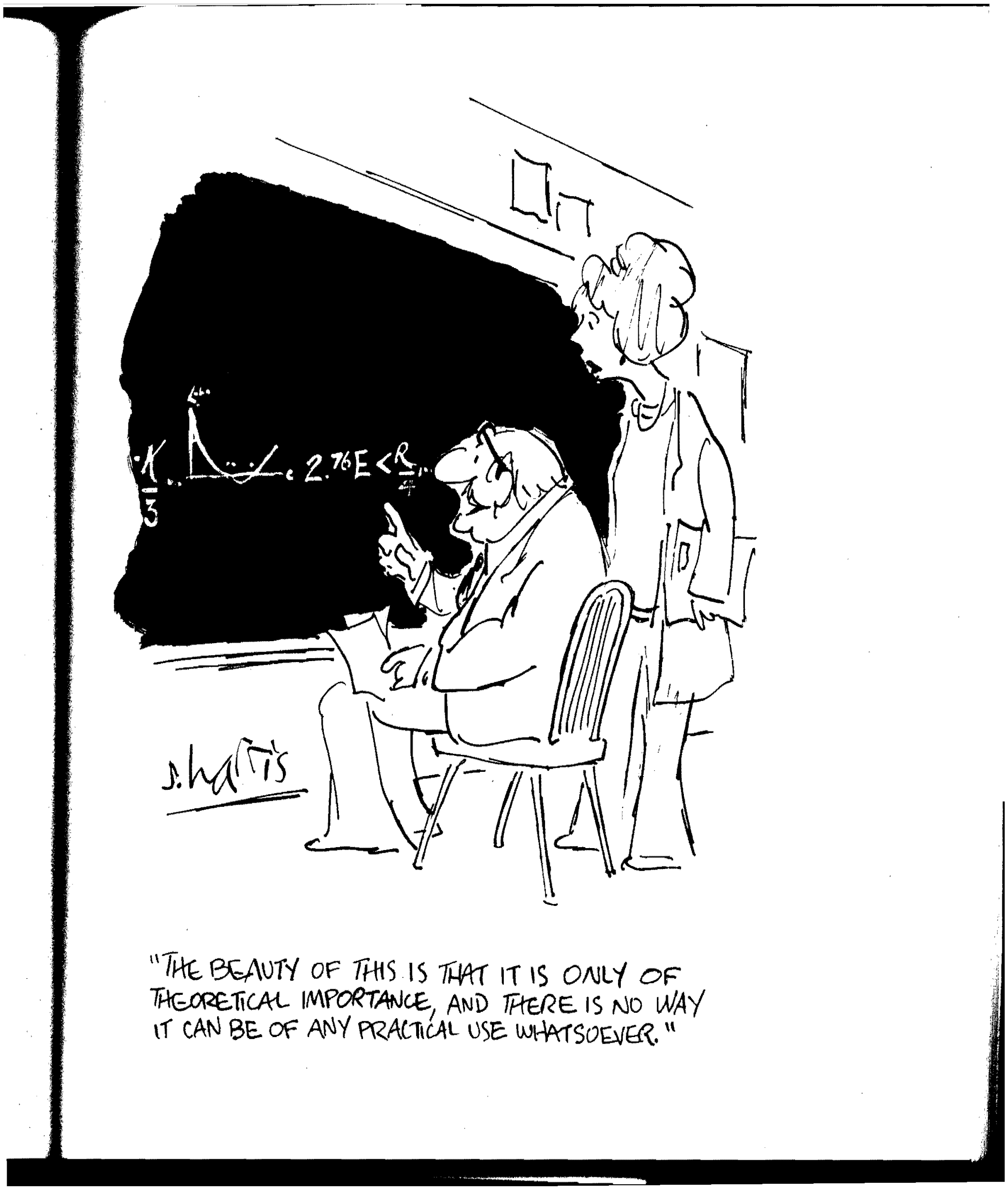


Figure 1.23: Beauty

1.14 The Four Fundamental Subspaces

Given a linear map $f\colon E \to F$ (where E and F are finite-dimensional), Proposition 1.33 revealed that the four spaces

$$\operatorname{Im} f,\ \operatorname{Im} f^{\top},\ \operatorname{Ker} f,\ \operatorname{Ker} f^{\top}$$

play a special role. They are often called the *fundamental subspaces* associated with f.

These spaces are related in an intimate manner, since Proposition 1.33 shows that

$$\begin{aligned} \operatorname{Ker} f &= (\operatorname{Im} f^{\top})^0 \\ \operatorname{Ker} f^{\top} &= (\operatorname{Im} f)^0, \end{aligned}$$

and Theorem 1.34 shows that

$$\operatorname{rk}(f) = \operatorname{rk}(f^{\top}).$$

It is instructive to translate these relations in terms of matrices (actually, certain linear algebra books make a big deal about this!).

If $\dim(E) = n$ and $\dim(F) = m$, given any basis $(u_1, \ldots, u_n)$ of E and a basis $(v_1, \ldots, v_m)$ of F, we know that f is represented by an $m \times n$ matrix $A = (a_{i\,j})$, where the jth column of A is equal to $f(u_j)$ over the basis $(v_1, \ldots, v_m)$.

Furthermore, the transpose map $f^\top$ is represented by the $n \times m$ matrix $A^\top$ (with respect to the dual bases).

Consequently, the four fundamental spaces

$$\mathrm{Im}\, f,\ \mathrm{Im}\, f^\top,\ \mathrm{Ker}\, f,\ \mathrm{Ker}\, f^\top$$

correspond to

(1) The *column space* of A, denoted by $\operatorname{Im} A$ or $\mathcal{R}(A)$; this is the subspace of $\mathbb{R}^m$ spanned by the columns of A, which corresponds to the image $\operatorname{Im} f$ of f.

(2) The *kernel* or *nullspace* of A, denoted by $\operatorname{Ker} A$ or $\mathcal{N}(A)$; this is the subspace of $\mathbb{R}^n$ consisting of all vectors $x \in \mathbb{R}^n$ such that $Ax = 0$.

(3) The *row space* of A, denoted by $\operatorname{Im} A^\top$ or $\mathcal{R}(A^\top)$; this is the subspace of $\mathbb{R}^n$ spanned by the rows of A, or equivalently, spanned by the columns of $A^\top$, which corresponds to the image $\operatorname{Im} f^\top$ of $f^\top$.

(4) The *left kernel* or *left nullspace* of A denoted by $\operatorname{Ker} A^\top$ or $\mathcal{N}(A^\top)$; this is the kernel (nullspace) of $A^\top$, the subspace of $\mathbb{R}^m$ consisting of all vectors $y \in \mathbb{R}^m$ such that $A^\top y = 0$, or equivalently, $y^\top A = 0$.

Recall that the dimension r of $\operatorname{Im} f$, which is also equal to the dimension of the column space $\operatorname{Im} A = \mathcal{R}(A)$, is the *rank* of A (and f).

Then, some our previous results can be reformulated as follows:

1. The column space $\mathcal{R}(A)$ of A has dimension r.
2. The nullspace $\mathcal{N}(A)$ of A has dimension $n - r$.
3. The row space $\mathcal{R}(A^\top)$ has dimension r.
4. The left nullspace $\mathcal{N}(A^\top)$ of A has dimension $m - r$.

The above statements constitute what Strang calls the *Fundamental Theorem of Linear Algebra, Part I* (see Strang [30]).

The two statements

$$\begin{aligned} \operatorname{Ker} f &= (\operatorname{Im} f^\top)^0 \\ \operatorname{Ker} f^\top &= (\operatorname{Im} f)^0 \end{aligned}$$

translate to

(1) The nullspace of A is the orthogonal of the row space of A.

(2) The left nullspace of A is the orthogonal of the column space of A.

The above statements constitute what Strang calls the *Fundamental Theorem of Linear Algebra, Part II* (see Strang [30]).

Since vectors are represented by column vectors and linear forms by row vectors (over a basis in E or F), a vector $x \in \mathbb{R}^n$ is orthogonal to a linear form y if

$$yx = 0.$$

Then, a vector $x \in \mathbb{R}^n$ is orthogonal to the row space of A iff x is orthogonal to every row of A, namely $Ax = 0$, which is equivalent to the fact that x belong to the nullspace of A.

Similarly, the column vector $y \in \mathbb{R}^m$ (representing a linear form over the dual basis of F^*) belongs to the nullspace of $A^\top$ iff $A^\top y = 0$, iff $y^\top A = 0$, which means that the linear form given by $y^\top$ (over the basis in F) is orthogonal to the column space of A.

Since (2) is equivalent to the fact that *the column space of A is equal to the orthogonal of the left nullspace of A*, we get the following criterion for the solvability of an equation of the form $Ax = b$:

The equation $Ax = b$ has a solution iff for all $y \in \mathbb{R}^m$, if $A^\top y = 0$, then $y^\top b = 0$.

Indeed, the condition on the right-hand side says that b is orthogonal to the left nullspace of A, that is, that b belongs to the column space of A.

This criterion can be cheaper to check that checking directly that b is spanned by the columns of A. For example, if we consider the system

$$\begin{aligned} x_1 - x_2 &= b_1 \\ x_2 - x_3 &= b_2 \\ x_3 - x_1 &= b_3 \end{aligned}$$

which, in matrix form, is written $Ax = b$ as below:

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

we see that the rows of the matrix A add up to 0.

In fact, it is easy to convince ourselves that the left nullspace of A is spanned by $y = (1, 1, 1)$, and so the system is solvable iff $y^\top b = 0$, namely

$$b_1 + b_2 + b_3 = 0.$$

Note that the above criterion can also be stated negatively as follows:

The equation $Ax = b$ has no solution iff there is some $y \in \mathbb{R}^m$ such that $A^\top y = 0$ and $y^\top b \neq 0$.

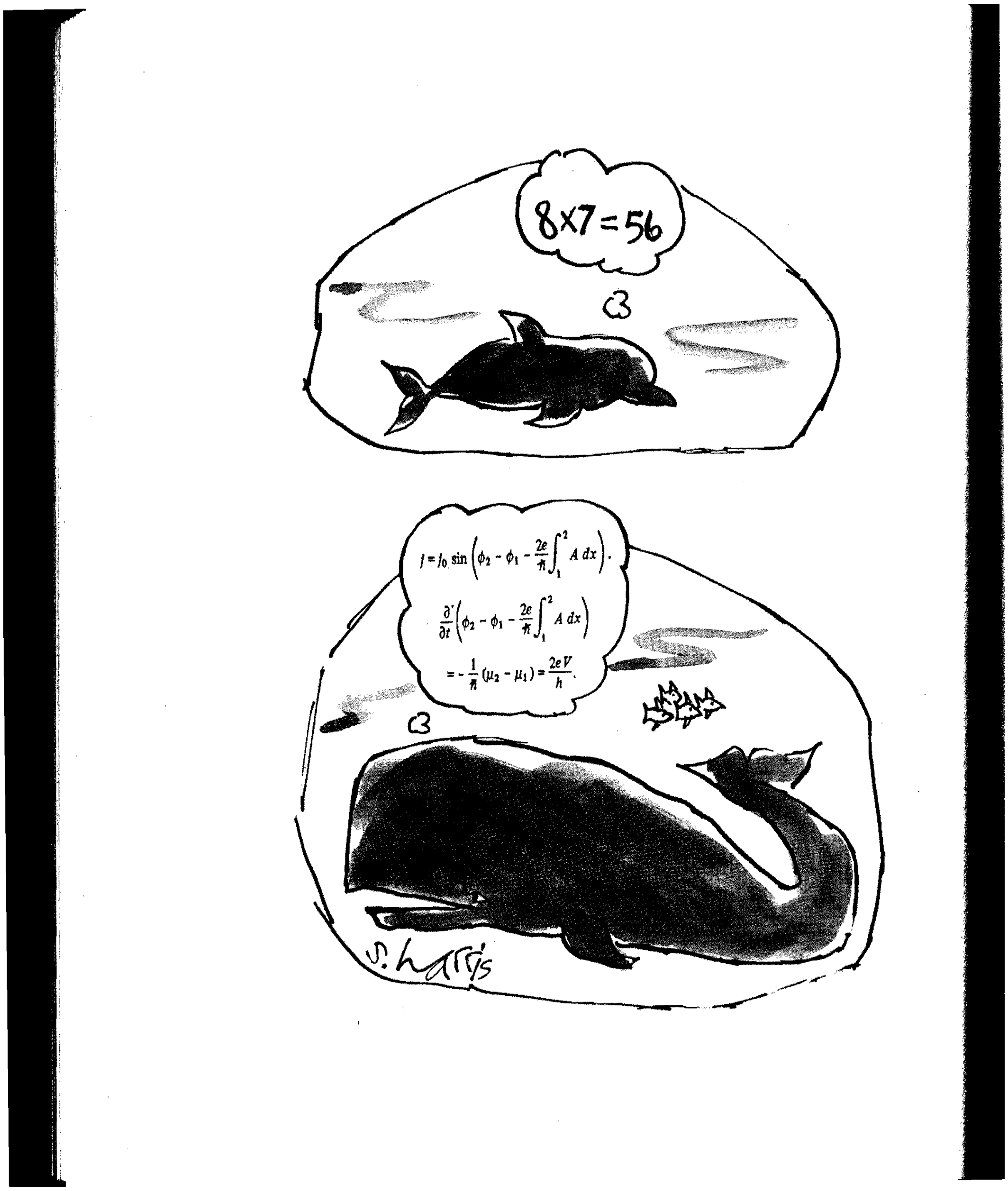


Figure 1.24: Brain Size?