Matrices and Vectors. . . in a Nutshell

AT Patera, M Yano

October 9, 2014

1 Preamble

Many phenomena and processes are well represented by linear mathematical models: linear relationships between the variables which describe the behavior of system. In many cases, for example related to the discretization of continuous media, the number of variables can be very large. Linear algebra provides the language and the theory with which (small and) large sets of linear equations can be readily manipulated and understood, and the foundation upon which effective computational approaches may be developed and implemented.

In this nutshell, we introduce the two objects which figure most prominently in linear algebra: vectors and matrices.

> We introduce operations on vectors: vector scaling; vector addition; the inner product; the norm. We present the triangle inequality, the Cauchy-Schwarz inequality, and we develop the concepts of orthogonality and orthonormality.
>
> We introduce operations on matrices: matrix scaling; matrix addition; matrix multiplication; the transpose and the transpose product rule; and matrix "division" (the inverse), briefly.
>
> We elaborate on matrix multiplication: non-commutation; the role of the identity matrix; the row and column interpretations of the matrix-vector product; operation counts.
>
> We develop the concept of linear combinations of vectors, the associated properties of linear independence and linear dependence, and the representation of linear combinations as a matrix-vector product.

Where possible, we emphasize simple geometric interpretations in two and three dimensions.

Prerequisites: secondary-school algebra and analytic geometry; some familiarity with low-dimensional vectors in the context of elementary mechanics.

2 Vectors and Matrices

Let us first introduce the primitive objects in linear algebra: vectors and matrices. A m-vector v consists of m real numbers — referred to as components, or elements, or entries —arranged in a column:[1]

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} ;$$

note that v_i refers to the i^{th} component of the vector. The other kind of vector is a row vector: a row n-vector v consists of n components now arranged in a row,

$$v = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} .$$

This column form is the default in linear algebra: "vector" absent a column or row adjective shall refer to a column vector.

We provide first a few examples of (column) vectors in $\mathbb{R}^3$:

$$v = \begin{pmatrix} 1 \\ 3 \\ 6 \end{pmatrix}, \quad u = \begin{pmatrix} \sqrt{3} \\ -7 \\ \pi \end{pmatrix}, \quad \text{and} \quad w = \begin{pmatrix} 9.1 \\ 7/3 \\ \sqrt{\pi} \end{pmatrix} .$$

To address a specific component of any of these vectors we write, for example, $v_1 = 1$, $u_1 = \sqrt{3}$, and $w_3 = \sqrt{\pi}$. Examples of row vectors in $\mathbb{R}^{1\times 4}$ are

$$v = \begin{pmatrix} 2 & -5 & \sqrt{2} & e \end{pmatrix} \quad \text{and} \quad u = \begin{pmatrix} -\sqrt{\pi} & 1 & 1 & 0 \end{pmatrix} .$$

Some of the components of these row vectors are $v_2 = -5$ and $u_4 = 0$.

A matrix A — an $m \times n$, or m by n, matrix — consists of m rows and n columns for a total of $m \cdot n$ entries (or elements),

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} ;$$

note A_{ij} refers the entry on the i^{th} row and j^{th} column. We may directly define A in terms of its entries as

$$A_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n .$$

[1] The concept of vectors readily extends to complex numbers, but we only consider real vectors in this nutshell.

Note that a column vector is a special case of a matrix with only one column, $n = 1$: a column m-vector is an $m \times 1$ matrix. Similarly, a row vector is a special case of a matrix with only one row, $m = 1$: a row n-vector is a $1 \times n$ matrix. Conversely, an $m \times n$ matrix can be viewed as m row n-vectors or n column m-vectors, as we discuss further below.

We provide a few examples of matrices:

$$A = \begin{pmatrix} 1 & \sqrt{3} \\ -4 & 9 \\ \pi & -3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 0 & 1 \\ -2 & 8 & 1 \\ 0 & 3 & 0 \end{pmatrix} .$$

The matrix A is a 3×2 matrix, and the matrix B is a 3×3 matrix. We can also address specific entries as, for example, $A_{12} = \sqrt{3}$, $A_{31} = \pi$, and $B_{32} = 3$.

While vectors and matrices may appear to be simply arrays of numbers, linear algebra defines a special set of rules for manipulation of these objects. One such operation is the transpose operation, denoted by superscript $(\cdot)^{\mathrm{T}}$. The transpose operator exchanges the rows and columns of the matrix: if $B = A^{\mathrm{T}}$ for A an $m \times n$ matrix, then

$$B_{ij} = A_{ji}, \quad 1 \le i \le n, \quad 1 \le j \le m .$$

Because the rows and columns of the matrix are exchanged, the dimensions of the matrix are also exchanged: if A is $m \times n$, then B is $n \times m$. If we exchange the rows and columns twice, then we return to the original matrix: $(A^{\mathrm{T}})^{\mathrm{T}} = A$. A *symmetric* matrix is a matrix A for which $A^{\mathrm{T}} = A$: a symmetric matrix is perforce a square matrix, $m = n$.

Let us consider a few examples of the transpose operation. A 3×2 matrix A yields an 2×3 transpose $B = A^{\mathrm{T}}$:

$$A = \begin{pmatrix} 1 & \sqrt{3} \\ -4 & 9 \\ \pi & -3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & -4 & \pi \\ \sqrt{3} & 9 & -3 \end{pmatrix} .$$

The rows and columns are exchanged: the first column of A is the first row of $B = A^{\mathrm{T}}$, and the second column of A is the second row of $B = A^{\mathrm{T}}$. In terms of individual entries, the indices are exchanged: $A_{31} = B_{13} = \pi$, and $A_{12} = B_{21} = \sqrt{3}$. We can also apply the transpose operator to a column vector (as a special case of a matrix): the column vector

$$v = \begin{pmatrix} \sqrt{3} \\ -7 \\ \pi \end{pmatrix},$$

yields as transpose

$$u = v^{\mathrm{T}} = \begin{pmatrix} \sqrt{3} & -7 & \pi \end{pmatrix}.$$

Note that the transpose of a column vector is a row vector, and the transpose of a row vector is a column vector.

We conclude with a remark on notation and visualization (or geometric interpretation). We can visualize a 1-vector $v = v_1$ as a point, or more precisely a ray, on the real-line: we locate the tail or root of the vector at 0, and then place the head — typically indicated as an arrow — of the vector at v_1 (the single entry of v). As a single coordinate suffices, we can interpret the real line as "1-space," also denoted $\mathbb{R}^1$. We visualize a 2-vector $v = \begin{pmatrix} v_1 & v_2 \end{pmatrix}^{\mathrm{T}}$ as a ray in the Cartesian plane: we place the tail of the vector at $(0, 0)$, and the head of the vector at (v_1, v_2). As two coordinates suffice, we can interpret the plane as "2-space," also denoted $\mathbb{R}^2$. We shall often take advantage of $\mathbb{R}^2$ — easily depicted on the page — to provide geometric interpretations of various concepts in linear algebra. We can readily extend these definitions and notation to higher dimensions, even if for $m > 3$ the visualization is difficult: an m-vector (or m-dimensional vector) v resides in m-space, which we may write more succinctly as $v \in \mathbb{R}^m$; the m elements of v serve to identify the m coordinates of the head of the vector. An $m \times n$ matrix A is often denoted, in an analogous fashion, as $A \in \mathbb{R}^{m \times n}$.

2.1 Vector Operations

2.1.1 Vector Scaling and Vector Addition

The first vector operation we consider is multiplication of a vector by a scalar, or *vector scaling*. Given an m-vector v and a scalar α, the operation $(u =)\, \alpha v$ yields an m-vector u with entries $u_i = \alpha v_i, i = 1, \ldots, m$. In words, to obtain $u = \alpha v$, we scale (multiply) each entry of v by α. Note that if $\alpha = -1$ then $\alpha u = -v$ for $-v$ the vector with entries $-v_i, 1 \leq i \leq n$.

The second operation we consider is vector addition. Given two m-vectors, v and w, the vector addition $(u =)\, v+w$ yields the vector m-vector u with entries $u_i = v_i + w_i, i = 1, \ldots, m$. In words, each entry of $u = v + w$ is the sum of the corresponding entries of v and w. We may also consider the addition of n vectors: each entry of the sum of n vectors is the sum of the corresponding entries of each of the n vectors. (Note vector addition is associative and commutative — we may group and order the summands as we wish.) Note we may only add vectors of the same dimension, m.

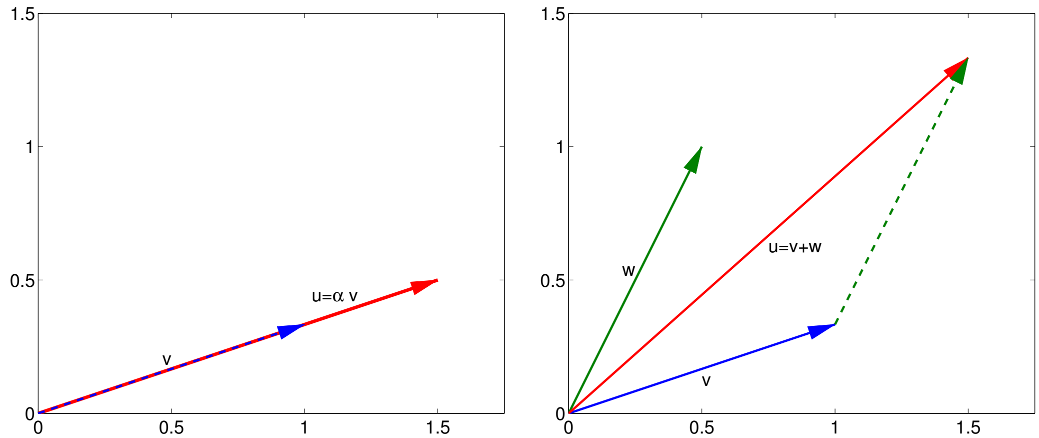
We now present a numerical example of each of these two fundamental vector operations. We consider $m = 2$ and

$$v = \begin{pmatrix} 1 \\ 1/3 \end{pmatrix}, \quad w = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}, \quad \text{and} \quad \alpha = \frac{3}{2} \ .$$

First, let us consider multiplication of the vector v by the scalar α, or vector scaling: the operation yields

$$u = \alpha v = \frac{3}{2} \begin{pmatrix} 1 \\ 1/3 \end{pmatrix} = \begin{pmatrix} 3/2 \\ 1/2 \end{pmatrix} .$$

Figure 1(a) illustrates the vector scaling process: to obtain the vector u we stretch the vector v by the factor of $3/2$; note u preserves the direction of v. Next, let us consider addition of



(a) scalar scaling (b) vector addition

Figure 1: Illustration of vector scaling and vector addition.

the two vectors v and w: vector addition. The operation yields

$$u = v + w = \begin{pmatrix} 1 \\ 1/3 \end{pmatrix} + \begin{pmatrix} 1/2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3/2 \\ 4/3 \end{pmatrix} .$$

Figure 1(b) illustrates the vector addition process: we translate (but do not rotate) w such that the tail of w coincides with the head of v; the vector from the tail of v to the head of the translated w then represents the sum of the two vectors, u. We can also think of u as the diagonal of the parallelogram generated by v and the translated w.

How do we combine the operations of vector scaling and vector addition? For $v \in \mathbb{R}^m$, $w \in \mathbb{R}^m$, and $\alpha \in \mathbb{R}$ (a scalar), the operation $(u =)\, v + \alpha w$ yields a vector $u \in \mathbb{R}^m$ with entries $u_i = v_i + \alpha w_i, i = 1, \ldots, m$. In short, we first apply the operation of vector scaling to obtain a vector $w' = \alpha w$; we then apply the operation of vector addition to obtain $u = v + w'$. For $\alpha = -1$ we of course arrive at vector subtraction.

CYAWTP 1. Let $v = \begin{pmatrix} 1 & 3 & 6 \end{pmatrix}^{\mathrm{T}}$, $w = \begin{pmatrix} 2 & -1 & 0 \end{pmatrix}^{\mathrm{T}}$, and $\alpha = 3$. Find $u = v + \alpha w$.

(Note we shall often define column vectors as the transpose ($^{\mathrm{T}}$) of a row vector purely for more economical use of space on the page.)

2.1.2 Inner Product and Norm

Another important vector operation is the inner product. This operation takes two m-vectors, v and w, and yields a scalar $(\beta =)\, v^{\mathrm{T}} w$, where

$$v^{\mathrm{T}} w \equiv \sum_{i=1}^{m} v_i w_i \; ; \tag{1}$$

it follows from this definition that $v^{\mathrm{T}}w = w^{\mathrm{T}}v$. The reason for the $v^{\mathrm{T}}w$ notation for the inner product — in particular, the appearance of the transpose operator — will become clear once we introduce the matrix-matrix multiplication rule. In any event, quite independent of notation, the inner product of two m-vectors, v and w, is uniquely defined by (1): we multiply corresponding components of v and w and sum the resulting m terms. The inner product is also sometimes referred to as the dot product (in particular in two and three dimensions), which is denoted as $\beta = v \cdot w$. Note we may only take the inner product of two vectors of the same dimension, m.

We can now define the 2-norm (which we shall subsequently refer to simply as the norm) of a vector. Given an m-vector v, the norm of v, denoted by $\|v\|_2$, is defined by

$$\|v\| \equiv \sqrt{v^{\mathrm{T}}v} = \sqrt{\sum_{i=1}^{m} v_i^2} \ .$$

Note that the norm of any vector v is a *non-negative* real number since $\|v\|^2$ is the sum of m squares. In two dimensions, $m = 2$, the norm of a vector v, $\sqrt{v_1^2 + v_2^2}$, is clearly the geometric *length* of v; we may thus interpret norm as the generalization of length (or distance) to any dimension. We know that the shortest distance between two points is a straight line. In m dimensions, this observation yields the *triangle inequality*: for any two m-vectors v and w,$\|v + w\| \leq \|v\| + \|w\|$; this result is clear from Figure 1(b).

CYAWTP 2. Consider the two vectors in $\mathbb{R}^3$, $v = \begin{pmatrix} 1 & 3 & 6 \end{pmatrix}^{\mathrm{T}}$ and $w = \begin{pmatrix} 2 & -1 & 0 \end{pmatrix}^{\mathrm{T}}$. Find the inner product $v^{\mathrm{T}}w$ and the norms $\|v\|$, $\|w\|$, and $\|v + w\|$. Verify that the triangle inequality is satisfied.

In two dimensions, the inner product (or "dot product") can be expressed as

$$v^{\mathrm{T}}w = \|v\|\|w\| \cos(\theta) \ , \tag{2}$$

where θ is the angle between v and w. We observe, since $|\cos(\theta)| \leq 1$, that

$$|v^{\mathrm{T}}w| = \|v\|\|w\||\cos(\theta)| \leq \|v\|\|w\| \ ;$$

this inequality, the famous Cauchy-Schwarz inequality, is in fact valid in any dimension m. We observe from (2) that the inner product is a measure of how well v and w *align* with each other: if $\theta = \pi/2$ the two vectors are perpendicular — maximally unaligned — and since $\cos(\pi/2) = 0$ then $v^{\mathrm{T}}w = 0$; if $\theta = 0$ the two vectors are colinear — maximally aligned — and since $\cos(0) = 1$ then $v^{\mathrm{T}}w = \|v\|\|w\|$. The Cauchy-Schwarz inequality implies that the magnitude of the inner product between two vectors is largest if the two vectors are colinear.

2.1.3 Orthogonality and Orthonormality

Our discussion of the inner product leads naturally to the concept of orthgonality. Two vectors $v \in \mathbb{R}^m$ and $w \in \mathbb{R}^m$ are said to be *orthogonal* (to each other) if

$$v^{\mathrm{T}}w = 0 \ .$$

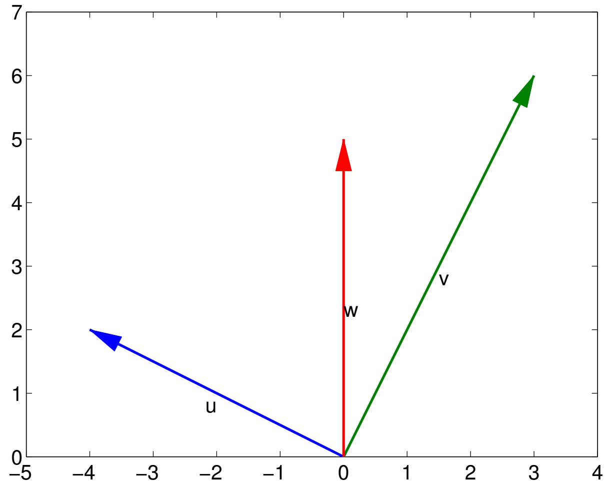


Figure 2: Illustration of orthogonality: $u^{\mathrm{T}}v = 0$.

We know from our discussion in two dimensions that "orthogonal" conforms with our usual geometric sense of "perpendicular." We shall further say that two vectors $v \in \mathbb{R}^m$ and $w \in \mathbb{R}^m$ are *orthonormal* (to each other) if

$$v^{\mathrm{T}}w = 0 \quad \text{and} \quad \|v\| = \|w\| = 1 \; ;$$

orthonormal vectors are orthogonal vectors which are furthermore of unit length (more precisely, unit norm).

We present an example. Let us consider the three vectors in $\mathbb{R}^2$,

$$u = \begin{pmatrix} -4 \\ 2 \end{pmatrix}, \quad v = \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \quad \text{and} \quad w = \begin{pmatrix} 0 \\ 5 \end{pmatrix} \, ,$$

and calculate the three inner products which can be formed from these vectors:

$$\begin{aligned} u^{\mathrm{T}}v &= -4 \cdot 3 + 2 \cdot 6 = 0 \; , \\ u^{\mathrm{T}}w &= -4 \cdot 0 + 2 \cdot 5 = 10 \; , \\ v^{\mathrm{T}}w &= 3 \cdot 0 + 6 \cdot 5 = 30 \; . \end{aligned}$$

We first note that $u^{\mathrm{T}}v = 0$, and hence the vectors u and v are orthogonal; however, the vectors u and v are not of norm unity, and hence the vectors u and m are not orthonormal. We further note that $u^{\mathrm{T}}w \neq 0$, and hence the vectors u and w are not orthogonal; similarly, $v^{\mathrm{T}}w \neq 0$, and hence the vectors v and w are not orthogonal. The vectors u, v, and w are presented in Figure 2; we confirm that u and v are perpendicular in the usual geometric sense.

CYAWTP 3. Show that the two vectors $u = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 & 1 \end{pmatrix}^{\mathrm{T}}$ and $v = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \end{pmatrix}^{\mathrm{T}}$ are orthogonal and furthermore orthonormal. Provide a sketch of u and v in the Cartesian plane.

2.1.4 Linear Combinations of Vectors

Finally, we consider the notion of a linear combination of vectors. Given a set of n m-vectors, $v^1 \in \mathbb{R}^m$, $v^2 \in \mathbb{R}^m, \ldots, v^n \in \mathbb{R}^m$, and an n–vector $\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}^{\mathrm{T}}$, the sum

$$(w =) \sum_{j=1}^{n} v^j \alpha_j$$

is denoted a *linear combination*. In words, the m-vector linear combination w is an α-weighted sum of the $v^j, 1 \leq j \leq n$.

We provide an example of a linear combination of vectors. We choose m — the dimension of the space in which our vectors reside — $= 2$, and n — the number of vectors we shall combine — $= 3$. We consider for our $v^j, 1 \leq j \leq n \equiv 3$ the 2-vectors $v^1 = \begin{pmatrix} -4 & 2 \end{pmatrix}^{\mathrm{T}}$, $v^2 = \begin{pmatrix} 3 & 6 \end{pmatrix}^{\mathrm{T}}$, and $v^3 = \begin{pmatrix} 0 & 5 \end{pmatrix}^{\mathrm{T}}$. We chose for our "weights" $\alpha_1 = 1$, $\alpha_2 = -2$, and $\alpha_3 = 3$, and hence $\alpha = \begin{pmatrix} 1 & -2 & 3 \end{pmatrix}^{\mathrm{T}}$. We thus obtain

$$\begin{aligned} w = \sum_{j=1}^{3} \alpha_j v^j &= 1 \cdot \begin{pmatrix} -4 \\ 2 \end{pmatrix} + (-2) \cdot \begin{pmatrix} 3 \\ 6 \end{pmatrix} + 3 \cdot \begin{pmatrix} 0 \\ 5 \end{pmatrix} \\ &= \begin{pmatrix} -4 \\ 2 \end{pmatrix} + \begin{pmatrix} -6 \\ -12 \end{pmatrix} + \begin{pmatrix} 0 \\ 15 \end{pmatrix} \\ &= \begin{pmatrix} -10 \\ 5 \end{pmatrix}. \end{aligned}$$

Note that the first line is the definition of a linear combination; the second line results from application (three times) of vector scaling; the third line results from application (two times) of vector addition.

We may now define the important concepts of linear independence (and linear dependence). We say that n m-vectors $v^j, 1 \leq j \leq n$, are *linearly **inde**pendent* if the condition

$$\sum_{j=1}^{n} v^j \alpha_j = 0 \quad \text{only if} \quad \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0 \tag{3}$$

is satisfied. We say the vectors v^j, $1 \leq j \leq n$, are *linearly **de**pendent* if the condition (3) is not satisfied. In words, a set of vectors is linearly dependent if there exists a non-zero n-vector α which yields a zero (m-vector) linear combination.

We provide an example of linearly independent vectors. Consider the $n = 3$ m-vectors v^1, v^2, and v^3 given by

$$v^1 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \quad v^2 = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}, \quad \text{and} \quad v^3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix};$$

note that $m = 3$. We now look for an n-vector $\alpha = (\ \alpha_1 \quad \alpha_2 \quad \alpha_3\ )^{\mathrm{T}}$ such that

$$v^1\alpha_1 + v^2\alpha_2 + v^3\alpha_3 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}\alpha_1 + \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}\alpha_2 + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\alpha_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} . \tag{4}$$

But application of our vector scaling and vector addition operations yields

$$\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}\alpha_1 + \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}\alpha_2 + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\alpha_3 = \begin{pmatrix} 2\alpha_1 \\ \alpha_3 \\ 3\alpha_1 \end{pmatrix} . \tag{5}$$

It follows from (4) and (5) that we wish to find α_1, α_2, and α_3 such that

$$\begin{pmatrix} 2\alpha_1 \\ \alpha_3 \\ 3\alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} . \tag{6}$$

We now examine this equation.

We first emphasize that two vectors are equal if and only if each component of the two vectors is equal. Hence the *vector equation* (6) is in fact m (scalar) equations — one equation for each component.[2] The first equation — equality of the first component — reads $2\alpha_1 = 0$, for which the unique solution for α_1 is 0; the second equation — equality of the second component — reads $\alpha_3 = 0$, for which the unique solution for α_3 is 0; the third equation — equality of the third component — reads $3\alpha_2 = 0$, for which the unique solution for α_2 is 0. We now return to our original quest: the *only* way in which we can satisfy (4) is to choose $\alpha = (\ 0\ 0\ 0\ )^{\mathrm{T}}$; it follows that (3) is *satisfied*, and hence the set of vectors v^1, v^2, and v^3 is *linearly independent*.

CYAWTP 4. Consider the $n = 3$ $(m =)2$-vectors $v^1 = (\ 1 \quad 0\ )^{\mathrm{T}}$, $v^2 = (\ 0 \quad 1\ )^{\mathrm{T}}$, and $v^3 = (\ 1 \quad 1\ )^{\mathrm{T}}$. Demonstrate that $v^1 + v^2 - v^3 = 0$. Are v^1, v^2, and v^3 linearly independent?

It can be shown that, in general, $n > m$ m-vectors $v^j, j = 1, \ldots, n$, *must* be linearly dependent; **CYAWTP 4** provides a numerical example. In can further be shown that any $n = m$ linearly independent m-vectors v^j, $j = 1, \ldots, n$, can be uniquely linearly combined to form any m-vector u; we may think of the v^j, $j = 1, \ldots, n$, as a — one possible set — of "basis" vectors in terms of which we can represent (any) member of m-space. The classical Cartesian unit vectors associated with coordinate directions are an example of basis vectors in 2-space ($m = 2$) or 3-space ($m = 3$).

[2] We have been, and will continue to be, rather sloppy in our notation. Many of our "equal" signs, $=$, would better be replaced by "defined as" or "equivalent" symbols, $\equiv$, or perhaps "assignment" operations, $\leftarrow$. In contrast, in (6), we encounter a *bona fide* equation: we wish to find α such that (6) is true.

3 Matrix Operations

Recall that a matrix $A \in \mathbb{R}^{m\times n}$ — an $m \times n$ matrix — consists of m rows and n columns for a total of $m \cdot n$ entries,

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} .$$

This matrix can be interpreted in a column-centric manner as a set of n column m-vectors; alternatively, the matrix can be interpreted in a row-centric manner as a set of m row n-vectors. The case in which $m = n$ is particularly important: a matrix A with $m = n$ — the same number of rows and columns — is denoted a *square* $(n \times n)$ matrix.

3.1 Matrix Scaling and Matrix Addition

The first matrix operation we consider is multiplication of a matrix by a scalar, or *matrix scaling*. Given a matrix $A \in \mathbb{R}^{m_1\times n_1}$ and a scalar α, the operation

$$(B =)\, \alpha A \ ,$$

yields a matrix $B \in \mathbb{R}^{m_1\times n_1}$ with entries

$$B_{ij} = \alpha A_{ij}, \quad i = 1, \dots, m_1,\ j = 1, \dots, n_1 \ .$$

In words, to obtain αA, we scale (multiply) each entry of A by α. Note that if $\alpha = -1$ then $\alpha A = -A$.

The second operation we consider is matrix addition. Given two matrices $A \in \mathbb{R}^{m_1\times n_1}$ and $B \in \mathbb{R}^{m_1\times n_1}$, the matrix addition

$$(C =)\, A + B \ ,$$

yields the matrix $C \in \mathbb{R}^{m_1\times n_1}$ with entries

$$C_{ij} = A_{ij} + B_{ij}, \quad i = 1, \dots, m_1,\ j = 1, \dots, n_1 \ .$$

We may also consider the addition of ℓ matrices: each entry of the sum of ℓ matrices is the sum of the corresponding entries of each of the ℓ matrices. Note matrix addition is associative and commutative — we may group and order the summands as we wish. Note we may only add matrices of the same dimensions.

We can easily combine matrix scaling and matrix addition. For $A \in \mathbb{R}^{m_1\times n_1}$, $B \in \mathbb{R}^{m_1\times n_1}$, and $\alpha \in \mathbb{R}$, the operation

$$(C =)\, A + \alpha B$$

yields a matrix $C \in \mathbb{R}^{m_1 \times n_1}$ with entries

$$C_{ij} = A_{ij} + \alpha B_{ij}, \quad i = 1, \ldots, m_1, \; j = 1, \ldots, n_1 \; .$$

We first apply the operation of matrix scaling to form a matrix $B' = \alpha B$; we then apply the operation of matrix addition to obtain $C = A + B'$.

CYAWTP 5. Let

$$A = \begin{pmatrix} 1 & \sqrt{3} \\ -4 & 9 \\ \pi & -3 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 2 \\ 2 & -3 \\ \pi & -4 \end{pmatrix}, \quad \text{and} \quad \alpha = 2 \; .$$

Construct the matrix $C = A + \alpha B$: provide all entries $C_{ij}, 1 \leq i \leq 3, 1 \leq j \leq 2$.

3.2 Matrix Multiplication

3.2.1 The Matrix-Matrix Product

We now proceed to matrix multiplication, which finally now introduces a deviation from our usual "scalar" intuition. Given two matrices $A \in \mathbb{R}^{m_1 \times n_1}$ and $B \in \mathbb{R}^{m_2 \times n_2}$ *which satisfy the condition* $n_1 = m_2$, the matrix-matrix product

$$(C =)\, AB$$

yields a matrix $C \in \mathbb{R}^{m_1 \times n_2}$ with entries

$$C_{ij} = \sum_{k=1}^{n_1} A_{ik} B_{kj}, \quad i = 1, \ldots, m_1, \; j = 1, \ldots, n_2 \; . \tag{7}$$

Because the summation is carried out over the second (column) index of A and the first (row) index of B, the number of columns of A, the first factor in the product, must match the number of rows of B, the second factor in our product: $n_1 = m_2$ *must* be satisfied for the product AB to be allowed. We also emphasize that C has dimensions $m_1 \times n_2$: C has the same number of rows as A, the first factor in our product, and the same number of columns as B, the second factor in our product.

We provide a detailed example of the mechanics of matrix multiplication (though we will later provide some more intuitive shortcuts in certain cases). Let us consider matrices $A \in \mathbb{R}^{3 \times 2}$ and $B \in \mathbb{R}^{2 \times 3}$ — note the number of columns of A matches the number of rows of B — with

$$A = \begin{pmatrix} 1 & 3 \\ -4 & 9 \\ 0 & -3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & 3 & -5 \\ 1 & 0 & -1 \end{pmatrix} \; .$$

The matrix-matrix product yields

$$C = AB = \begin{pmatrix} 1 & 3 \\ -4 & 9 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} 2 & 3 & -5 \\ 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 5 & 3 & -8 \\ 1 & -12 & 11 \\ -3 & 0 & 3 \end{pmatrix} . \tag{8}$$

Note that C has dimensions 3×3: the number of rows is inherited from A, the first factor in the product; the number of columns is inherited from B, the second factor in the product. We now expand out each entry to illustrate the summation over the "inner" index:

$$\begin{aligned}
C_{11} &= \sum_{k=1}^{2} A_{1k} B_{k1} = A_{11}B_{11} + A_{12}B_{21} = 1 \cdot 2 + 3 \cdot 1 = 5 \\
C_{12} &= \sum_{k=1}^{2} A_{1k} B_{k2} = A_{11}B_{12} + A_{12}B_{22} = 1 \cdot 3 + 3 \cdot 0 = 3 \\
C_{13} &= \sum_{k=1}^{2} A_{1k} B_{k3} = A_{11}B_{13} + A_{12}B_{23} = 1 \cdot (-5) + 3 \cdot (-1) = -8 \\
C_{21} &= \sum_{k=1}^{2} A_{2k} B_{k1} = A_{21}B_{11} + A_{22}B_{21} = (-4) \cdot 2 + 9 \cdot 1 = 1 \\
&\vdots \\
C_{33} &= \sum_{k=1}^{2} A_{3k} B_{k3} = A_{31}B_{13} + A_{32}B_{23} = 0 \cdot (-5) + (-3) \cdot (-1) = 3 \ .
\end{aligned}$$

We note that there will be $m_1 n_2 = 3 \cdot 3 = 9$ entries, each of which requires $2m_2 - 1 = 2n_1 - 1 = 3$ FLOPs. We will discuss operation counts more generally later in this nutshell.

CYAWTP 6. Consider the matrices A and B defined in (8). Confirm that the product $D = BA$ is allowed — satisfies our "matching inner index" requirement. Identify the dimensions of D. Finally, evaluate the entries of D.

The inner product of two vectors can be considered as a special case of the matrix-matrix product. In particular, for v an m-vector, and hence an $m \times 1$ matrix, and w an m-vector, and hence an $m\times$ 1 matrix, the product $v^{\mathrm{T}}w$ is the product of a 1 $\times m$ matrix and an $m\times$ 1 matrix. We thus identify in (7) $n_1 = 1$, $m_1 = m$ as the dimensions of the first factor in our product, and $n_2 = m, m_2 = 1$ as the dimensions of the second factor in our product. We can then conclude that (i) the matrix product $v^{\mathrm{T}}w$ is indeed allowed, since $m_1 = n_2$, and (ii) the matrix product $v^{\mathrm{T}}w$ will be of dimension $n_1 \times m_2 = 1 \times 1$ — a scalar, as desired.

CYAWTP 7. Let

$$v = \begin{pmatrix} 1 \\ 3 \\ 6 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} -2 \\ 0 \\ 4 \end{pmatrix} .$$

Calculate $v^{\mathrm{T}}w$, the inner product of v and w. Calculate the entries of wv^{T}, known as the "outer product" of w and v.

We have seen in the above examples that, in general, $AB \neq BA$. In many cases BA might not be allowed even if AB is allowed (consider $A \in \mathbb{R}^{3\times 2}$ and $B \in \mathbb{R}^{2\times 1}$). And even if AB and BA are both allowed, these two products might be of different dimensions — and hence patently different (consider our inner product and outer product example). And even if A and B are both square and of the same dimensions, such that AB and BA are both allowed and both of the same dimensions, $AB = BA$ only for matrices which satisfy a very special property. In short, matrix multiplication is not commutative in general: order matters. However, the matrix-matrix product *is* associative, $ABC = A(BC) = (AB)C$, and also distributive $(A+B)C = AC+BC$; these results are readily proven from the definition of the matrix product.

We close this section with an identity related to the matrix-matrix product. Given any two matrices $A \in \mathbb{R}^{n_1\times m_1}$ and $B \in \mathbb{R}^{n_2\times m_1}$ for which $m_1 = n_2$ (and hence the product AB is allowed),

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}} \ . \tag{9}$$

In words, the tranpose of the product is the product of the transposes — but evaluated in the "reverse" order.

CYAWTP 8. Consider the product $p = v^{\mathrm{T}}A^{\mathrm{T}}Av$ for v an n-vector and A an $m\times n$ matrix. Confirm that the necessary matrix multiplications are allowed. Show that p is a scalar. Demonstrate that $p = \|Av\|^2$ and hence always non-negative.

3.2.2 The Identity Matrix

We know that in the scalar case there is a unique real number — 1 — such that for any scalar a, $a = 1\cdot a$. In the case of matrices, the *identity* matrix, I^ℓ, plays a similar role. The identity matrix I^ℓ (note we may omit the superscript ℓ if the dimension is "understood") is an $\ell\times\ell$ square matrix with ones on the main diagonal and zeros elsewhere:

$$I^\ell_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \ .$$

Identity matrices in $\mathbb{R}^1$, $\mathbb{R}^2$, and $\mathbb{R}^3$ are given by

$$I^1 = \begin{pmatrix} 1 \end{pmatrix}, \quad I^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad I^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We see from (7) that, for any $m\times n$ matrix A,

$$(I^m A)_{ij} = \sum_{k=1}^{m} I^m_{ik} A_{kj} = A_{ij} \ ,$$

since for a given i only for $k = i$ is I^m_{ik} nonzero (and unity). We thus conclude that $I^m A = A$. We may also demonstrate, in the same fashion, that $AI^n = A$. Hence, as advertised, I^ℓ is the ℓ-dimensional version of "one."

3.2.3 Interpretations of the Matrix-Vector Product

Let us consider a special case of the matrix-matrix product $C = AB$: the matrix-vector product. In this case the first factor, A, is a general $m \times n$ matrix, whereas the second factor, B, is a (column) n-vector, hence an $n \times 1$ matrix. We first note that AB is allowed: the number of columns of A matches the number of rows or B — n. We next note that C is an $m \times 1$ matrix — hence a (column) m-vector: C inherits the m rows of A and the 1 column of B. Hence $C = AB$ in our particular case reduces to

$$v = Aw \ . \tag{10}$$

We now elaborate on the rather terse description (10).

We first recall that two vectors are equal if an only if each component of the two vectors is equal, and hence $v = Aw$ of (10) is actually m equations:

$$v_i = \sum_{k=1}^{n} A_{ik} w_k \ , \quad 1 \le i \le m \ ; \tag{11}$$

note that (11) is just a special case of (7). We may state (11) more explicitly as

$$v_1 \ = \ (Aw)_1 = \sum_{k=1}^{n} A_{1k} w_k \ , \tag{12}$$

$$v_2 \ = \ (Aw)_2 = \sum_{k=1}^{n} A_{2k} w_k \ , \tag{13}$$

$$\vdots \tag{14}$$

$$v_m = (Aw)_m = \sum_{k=1}^{n} A_{mk} w_k \ , \tag{15}$$

where $(Aw)_i$ refers to the i^{th} component of the *vector* Aw. Finally, and most explicitly, we can expand out the summation to obtain

$$\left.\begin{array}{lcl} v_1 & = & A_{11} w_1 + A_{12} w_2 + \cdots + A_{1n} w_n \\ v_2 & = & A_{21} w_1 + A_{22} w_2 + \cdots + A_{2n} w_n \\ & \vdots & \\ v_m & = & A_{m1} w_1 + A_{m2} w_2 + \cdots + A_{mn} w_n \end{array}\right\} \ . \tag{16}$$

We now consider two different interpretations of this matrix-vector product.

The first interpretation is the "row" interpretation. Here we consider the matrix-vector multiplication as a series of inner products. In particular, we interpret v_i as the inner product of the i^{th} row of A and our vector w: the vector v is evaluated "entry by entry" in the sense that

$$v_i = \begin{pmatrix} A_{i1} & A_{i2} & \cdots & A_{in} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}, \quad i = 1, \ldots, m \ .$$

This is the more traditional view of matrix multiplication, but oftentimes not the most englightening.

The second interpretation of the matrix-vector product is the "column" interpretation. Here we consider the matrix-vector multiplication as a w-weighted sum of the n column vectors which comprise A. In particular, from the definition of vector scaling and vector addition we recognize that

$$v = \begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} w_1 + \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} w_2 + \cdots + \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix} w_n \ .$$

exactly replicates (16). We thus observe that $v = Aw$ is a succinct fashion in which to express a linear combination of vectors as described in Section 2.1.4: the columns of A provide our n m-vectors, and the elements of w provide the coefficients.

We consider a particular matrix-vector product from both the row and column perspective. Our matrix A and vector w are given by

$$A \equiv \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad w \equiv \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \ .$$

First the row interpretation:

$$v = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 2 & 1 \end{pmatrix}^{\mathrm{T}} \\ \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 2 & 1 \end{pmatrix}^{\mathrm{T}} \\ \begin{pmatrix} 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 2 & 1 \end{pmatrix}^{\mathrm{T}} \\ \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 & 1 \end{pmatrix}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 0 \\ 1 \end{pmatrix} .$$

Next, the column interpretation:

$$v = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = 3 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 0 \\ 1 \end{pmatrix} .$$

Clearly, the outcome will be the same quite independent of how we choose to slice the operations.

CYAWTP 9. We introduce an $m = 10000 \times n = 5000$ matrix A the entries of which are all non-zero. We further introduce a column 5000-vector w, the entries of which are also all non-zero. We now consider the matrix-vector product $v = Aw$. How many FLOPs are required to calculate v_3 — the third entry, and only the third entry, of the vector v?

CYAWTP 10. We introduce an $m = 10000 \times n = 5000$ matrix A the entries of which are all non-zero. We further introduce a column 5000-vector w which is unity at entry $i = 3$ and zero at all other entries. Now consider the matrix-vector product $v = Aw$. How many FLOPs are required to evaluate $v = Aw$ — *all* entries of v?

3.2.4 Operation Counts

The matrix-matrix product is ubiquitous in scientific computation, with much effort expended to optimize the performance of this operation on modern computers. Let us now count the number of additions and multiplications — the FLOPs — required to calculate this product. (We recall that FLOPs is an acronym for *FL*oating point *OP*eration*s*.) We consider the multiplication of an $m_1 \times n_1$ matrix A and an $m_2 \times n_2$ matrix B to form an $m_1 \times n_2$ matrix C: $C = AB$. We of course assume that $n_1 = m_2$ such that the multiplication AB is allowed.

We know from (7), which we recall here, that $C = AB$ is given by

$$C_{ij} = \sum_{k=1}^{n_1} A_{ik} B_{kj}, \quad i = 1, \ldots, m_1,\ j = 1, \ldots, n_2 \ .$$

The computation of any entry C_{ij} requires n_1 multiplications and n_1 additions (or perhaps $n_1 - 1$, but our interest is in large n_1), hence $2n_1$ FLOPs in total. There are $m_1 n_2$ entries in C, and thus the total operation count for evaluation of the entire matrix C is $2m_1 n_1 n_2$ FLOPs.

In order to better understand the general result we present in Table 3.2.4 the operation counts for several special but important and ubiquitous cases: the inner product; the matrix-vector product for a square matrix; and the matrix-matrix product for two square matrices. Since in the latter two cases our matrix is square we can summarize our results in terms of n, which we may interpret as the "size" of the problems.

Operation	Dimensions	Operation Count
inner product	$n_1 = m_1 = n$, $m_1 = n_2 = 1$	$2n$
matrix-vector product	$m_1 = n_1 = m_2 = n$, $n_2 = 1$	$2n^2$
matrix-matrix product	$m_1 = n_1 = m_2 = n_2 = n$	$2n^3$

Table 3.2.4. Operation counts for particular cases of the matrix-matrix product.

We note the increase in the exponent as we proceed from the inner product to the matrix-vector product and finally the matrix-matrix product. In particular, we observe that the operation count for the multiplication of two $n \times n$ matrices increases rapidly with n: if we double n, the number of operations increases by a factor of eight. However, we should emphasize that the operation counts of Table 3.2.4 assumes that our matrices are *full* — that all (or most of the) entries of the factors are non-zero. In actual practice, many matrices which arise from models of physical systems are *sparse* — there are only a few non-zero entries — in which case, for properly implementations, the operation counts can be greatly reduced.

3.3 Matrix "Division": The Inverse of a Matrix (Briefly)

We have now considered matrix multiplication in some detail. Can we also define matrix "division"? To begin, let us revert to the scalar case. If b is a scalar, and a is a non-zero scalar, we know that we can divide by a to form $x = b/a$. We may write this more suggestively as $x = a^{-1}b$ since of course $a^{-1} = 1/a$. Note we may define a^{-1} as that number y such that $ya = 1$. Hence we may view division as, first, the formation of the inverse, a^{-1}, and second, multiplication by the inverse, $x = a^{-1}b$. It is important to note that we can perform division only if a is non-zero.

We can now proceed to the matrix case "by analogy." We shall consider a square matrix A of dimension $n \times n$ since in fact a proper inverse exists only for a square matrix. Hence, given an $n \times n$ matrix A, can we find an inverse matrix A^{-1} such that $A^{-1}A = I^n$? We recall that I^n is the $n \times n$ identity matrix — the n-dimensional version of "one" — and hence $A^{-1}A = I^n$ is equivalent in the scalar case to $a^{-1}a = 1$. The answer is yes, if and only if the n columns of A are linearly independent; if the latter is satisfied, we say that A is *invertible*. In the case in which A is invertible, the inverse matrix is unique, and furthermore not only do we obtain $A^{-1}A = I^n$ but also $AA^{-1} = I^n$.

As a concrete example, we consider the case of a 2×2 matrix A which we write as

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

If the columns are to be independent we must have $a/b \neq c/d$ or $(ad)/(bc) \neq 1$ or $ad - bc \neq 0$. The inverse of A is then given by

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

Note that this inverse is only defined if $ad - bc \neq 0$. It is a simple matter to show by explicit matrix multiplication that $A^{-1}A = AA^{-1} = I^2$, as desired.

We present here one application of the inverse. We have considered previously the matrix-vector product, $v = Aw$, which we might characterize as a forward (or explicit) problem: given w, evaluate v. We can contrast this forward problem to the inverse (or implicit) problem, $Aw = v$: given v, find — "solve for" — w. This inverse problem has another, much more common name: solution of n equations in n unknowns; the n equations are associated with the n rows of A and v, and the n unknowns are the n elements of w. We can visualize the latter as (16) with (by convention) the left-hand and right-hand sides exchanged.

We now consider this inverse problem in slightly more detail. In particular, we first assume that A is a square $n \times n$ matrix with n independent columns such that A^{-1} exists. We can then multiply the equation $Aw = v$ through by A^{-1} to obtain an explicit expression for w: $w = A^{-1}v$. We note that the latter, once A^{-1} is formed, is simply a matrix multiplication. The analogy with the scalar case is direct: solution of $ax = b$ is given by $x = a^{-1}b$ $(= b/a)$. In actual fact, formation of the inverse matrix followed by matrix multiplication is typically a *very poor*, and often a disastrous, way to solve (computationally) a set of n equations in n unknowns. However, for n very small, $A^{-1}v$ can be convenient; in any event, it is important to recognize the connection between the inverse of A and the solution of linear systems of equations associated with A.

CYAWTP 11. We consider the matrix

$$A \equiv \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} .$$

Find A^{-1} and confirm by matrix multiplication that $A^{-1}A = AA^{-1} = I^2$, where I^2 is the 2×2 identity matrix. Now introduce the vector $v = (\begin{array}{cc} 2 & 3 \end{array})^{\mathrm{T}}$. Find $w = A^{-1}v$ and confirm that $v - Aw = 0$.