Matrices and Linear Transformations

(2/25/19)

Alex Nita

Abstract

Modern mathematics sets for itself the twin goals of *completeness* and *comprehensiveness* in its descriptive work. This can be understood as follows: First, the way in which math *characterizes* a concept should be complete, otherwise the concept is poo-pooed as underdeveloped, unrigorous, etc. This naturally demands a concept's *properties* be made *explicit*. But that isn't enough. The rub of the issue is that a concept's properties may all be made explicit and yet leave room for *uncertainty* as to the concept's *nature*, somehow. It's a question of *definition*. Do the properties *define* the concept, in the sense of *fence off completely* from any possible *misunderstanding*, or do they not? The avowed goal in math is to reach a point of development, in whatever area its attention may focus, where the *properties themselves completely characterize*, in the sense of *define* or *determine the meaning of*, the concept. This is what makes math such a reliable language, and so eminently useful in science, which needs a tight, linguistically pure and true, conceptual representation. It also makes computers and information technology possible, for what we want most there is for our machines to execute our commands faithfully, without mistakes, errors or halts. Of course, there is more to completeness than definition, because there is also the issue of *provability*, but let us not worry about that—what's important for us now is to understand where definitions fit into the scheme of things.

We will see this more clearly in the first section, on the **defining properties of matrices**, namely the ***algebraic properties*** of *matrix addition, scalar multiplication, matrix multiplication*, etc. Each of these is in fact a definition, because properties define mathematical concepts. Together they *characterize* the concept of *what it means to be a matrix*.

As to the second point, comprehensiveness, math seeks, additionally, to *fully* characterize a concept. For example, the concept of a matrix may be clearly and unambiguously defined in terms of a list of properties, but that doesn't mean we thereby *fully understand* matrices. To say we understand a concept, we need to *fully* map out its territories, and *thoroughly* elucidated its *nature*. To feel *modernly confident*, that is—beyond the reproach of even the most practical man—in saying we *understand* a concept, we should say it *convincingly*, *in terms of some other, more familiar, concepts*. Matrices, for example, are understood in terms of the algebraic properties of the *real numbers*, and the real numbers are more familiar. But this is not all: we want a comprehensive road map, a *complete classification* of the ins and outs of the concept (in terms of the more familiar). Can we classify all matrices into clumps—symmetric, skew-symmetric, diagonalizable, etc.? We'll try.

All of this, plus one more, is needed to satisfy a real, bonifide, modern mathematician: *categories*. Nowadays everything must be couched in categorical language, ever since Eilenberg and MacLane invented category theory in the 1940s (the same MacLane whose book *Mathematics: Form and Function*, I gave as supplementary reading!). Categories are a sort of meta-definition. They make a

conceptual distinction between two *types of concepts*, the (mathematical) *objects* versus the *relations* between the objects (the *functions*). The objects in linear algebra are the $\mathbb{R}^n$s, or more generally the vector spaces V, while *the functions are the $m \times n$ matrices*, or more generally *linear transformations*.

Modern mathematics takes as its guiding principle that ***relations between objects are more important than the objects themselves***, so it is no surprise that the vast majority of this class is devoted to matrices and linear transformations, relegating the $\mathbb{R}^n$s to supporting roles and occasional cameos. But those $\mathbb{R}^n$s never quite disappear, for without them we could never produce **coordinates**, and as we will see, coordinates become *the major thing to understand* in relation to matrices.

In this section, we begin a study of the definitions of the *functions* in the vector space category: matrices and linear transformation. The objects, the $\mathbb{R}^n$s, were defined and briefly studied in the notes 'Points and Vectors in $\mathbb{R}^n$.' We'll follow this with a brief survey of the immediate implications of the basic definitions, then follow that with a closer analysis of the major *types* of matrices and linear transformations.

1 Definition of Matrices

Definition 1.1 Let us define a **real** $m \times n$ **matrix** as an array of mn real numbers a_{ij} consisting of m rows and n columns,

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

The **indices** i and j in the expression a_{ij} indicate, respectively, the **row number**, i, and the **column number**, j. Thus, element a_{ij} sits in the ith row and the jth column. For example, $a_{2,3}$ is the real number in the second row, third column of A. The word "real" in the phrase "real matrix" refers to the fact that these entries a_{ij} are *real numbers*.

We could analogously define complex matrices, quaternionic matrices, integer matrices, etc., where in each case a_{ij} would be a complex number, a quaternionic number, an integer, etc. respectively. Towards the end of this semester we will have a chance to look at the complex case. ■

Example 1.2 The following is a real 3×2 matrix:

$$A = \begin{pmatrix} 1 & -1 \\ 2 & -4 \\ 0 & \sqrt{2} \end{pmatrix}, \qquad \text{And observe:} \quad \begin{matrix} a_{11} = 1, & a_{12} = -1 \\ a_{21} = 2, & a_{22} = -4 \\ a_{31} = 0, & a_{32} = \sqrt{2} \end{matrix}$$

while the following is a real 1×4 matrix:

$$B = \begin{pmatrix} -1 & 2 & 45 & -\frac{1}{\sqrt{3}} \end{pmatrix}, \quad \text{where} \quad \begin{matrix} b_{11} = -1, & b_{12} = 2, \\ b_{13} = 45, & b_{14} = -\frac{1}{\sqrt{3}} \end{matrix}$$ ■

Matrices, like vectors, are not mere objects. First of all, they have *structure—algebraic structure*. We can add and scale matrices in exactly the same way as we do vectors, *elementwise*. Further, we can *multiply* matrices whose dimensions are compatible. We explain all this now.

Definition 1.3 We can define **addition** of two matrices A and B of the same dimension $m \times n$ as we did with vectors, componentwise:

$$A + B = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

We define **scalar multiplication** of an $m \times n$ real matrix A by a real number a:

$$aA = a \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} aa_{11} & \cdots & aa_{1n} \\ \vdots & \ddots & \vdots \\ aa_{m1} & \cdots & aa_{mn} \end{pmatrix}$$ ■

Example 1.4 For example, here a sum of 3×2 matrices:

$$\begin{pmatrix} 1 & -1 \\ 2 & -4 \\ 0 & \sqrt{2} \end{pmatrix} + \begin{pmatrix} 20 & 9 \\ 5 & 3 \\ \sqrt{5} & 2 \end{pmatrix} = \begin{pmatrix} 1+20 & -1+9 \\ 2+5 & -4+3 \\ 0+\sqrt{5} & \sqrt{2}+2 \end{pmatrix} = \boxed{\begin{pmatrix} 21 & 8 \\ 7 & -1 \\ \sqrt{5} & 2+\sqrt{2} \end{pmatrix}}$$

and here is a scalar multiple of a 3×2 matrix.

$$5 \begin{pmatrix} 1 & -1 \\ 2 & -4 \\ 0 & \sqrt{2} \end{pmatrix} = \begin{pmatrix} 5 \cdot 1 & 5 \cdot (-1) \\ 5 \cdot 2 & 5 \cdot (-4) \\ 5 \cdot 0 & 5 \cdot \sqrt{2} \end{pmatrix} = \boxed{\begin{pmatrix} 5 & -5 \\ 10 & -20 \\ 0 & 5\sqrt{2} \end{pmatrix}}$$ ■

Example 1.5 $(-2) \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} (-2)1 & (-2)2 & (-2)3 \\ (-2)4 & (-2)5 & (-2)6 \end{pmatrix} = \boxed{\begin{pmatrix} -2 & -4 & -6 \\ -8 & -10 & -12 \end{pmatrix}}$ ■

Remark 1.6 Consider the future import of these two algebraic properties: They turn matrices into *vectors in a vector space (namely $M_{m,n}(\mathbb{R})$)*, as we will see later. Happy coincidence or calculated design choice? ■

Notation 1.7 *Let us introduce notation for the* ***set of all*** *$m \times n$* ***real matrices****,*

$$M_{m,n}(\mathbb{R}) = \left\{ A \;\middle|\; A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \right\}$$

We will often write

$$A \in M_{m,n}(\mathbb{R})$$

instead of saying *"A is an $m \times n$ matrix." Also,* we consider $M_{m,n}(\mathbb{R})$ to be a set equipped with algebraic structure, *which is how mathematicians look at these things.* ■

In order to carry out matrix multiplication, as well as for other purposes, such as row-reduction, it is convenient to consider the *rows* and *columns* of a matrix A. Much of our analysis of matrices and systems of equations will hinge on a correct understanding of the behavior of these rows and columns.

Notation 1.8 *Let $A \in M_{m,n}(\mathbb{R})$. In terms of its elements,*

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

In this class we will ***denote the rows of*** *A by $\vec{A}_i$:*

$$\begin{aligned} \vec{A}_1 &= \langle a_{11}, a_{12}, \dots, a_{1n} \rangle \\ \vec{A}_2 &= \langle a_{21}, a_{22}, \dots, a_{2n} \rangle \\ &\vdots \\ \vec{A}_m &= \langle a_{m1}, a_{m2}, \dots, a_{mn} \rangle \end{aligned}$$

and we ***denote the columns of*** *A by $\mathbf{a}_j$:*

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} a_{12} \\ \vdots \\ a_{m2} \end{pmatrix}, \quad \dots, \quad \mathbf{a}_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

I will stick to this notation in this class, but beware that my notation for rows is non-standard. There is no standard notation for rows, so I just invented one. ■

Example 1.9 Consider the matrix

$$A = \begin{pmatrix} 1 & -1 \\ 2 & -4 \\ 0 & \sqrt{2} \end{pmatrix} \in M_{3,2}(\mathbb{R})$$

from Example 1.2 above. Then it's row and column vectors are

$$\begin{aligned} \vec{A}_1 &= \langle 1, -1 \rangle, \\ \vec{A}_2 &= \langle 2, -4 \rangle, \\ \vec{A}_3 &= \langle 0, \sqrt{2} \rangle, \end{aligned} \qquad \text{and} \qquad \mathbf{a}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} -1 \\ -4 \\ \sqrt{2} \end{pmatrix}$$

■

We'll define matrix multiplication in two ways, one of which is easier to understand if, first, we do a simpler product: *of a matrix A with a column vector* $\mathbf{x}$. Anyway, this is handy to have when we consider systems of equations, because with matrix notation they can be written in the form $A\mathbf{x} = \mathbf{b}$. More on that later.

Definition 1.10 Let $A \in M_{m,n}(\mathbb{R})$ and let $\mathbf{x} = \langle x_1, \ldots, x_n \rangle \in \mathbb{R}^n$. *Note the number n:* it shows up in the row-size of A as well as in the size of $\mathbf{x}$, *and it has to be that way* for this to work.[1] We define **multiplication of the matrix A with the column vector $\mathbf{x}$** in two equivalent ways: *the first* in terms of the columns of A,

$$A\mathbf{x} = \begin{pmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \overset{\text{def 1}}{=} x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

Now, since we know the meaning of $x_j\mathbf{a}_j$,

$$x_j\mathbf{a}_j = x_i \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = \begin{pmatrix} x_j a_{1j} \\ \vdots \\ x_i a_{mj} \end{pmatrix}$$

for all $j = 1, \ldots, n$, and we also know how to add vectors, this definition reduces to the *second definition*, in terms of the rows of A:

$$A\mathbf{x} = \begin{pmatrix} -\vec{A}_1- \\ -\vec{A}_2- \\ \vdots \\ -\vec{A}_m- \end{pmatrix} \mathbf{x} \overset{\text{def 2}}{=} \begin{pmatrix} \vec{A}_1 \cdot \mathbf{x} \\ \vec{A}_2 \cdot \mathbf{x} \\ \vdots \\ \vec{A}_m \cdot \mathbf{x} \end{pmatrix}$$

as follows:

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \begin{pmatrix} x_1 a_{11} + \cdots + x_n a_{1n} \\ \vdots \\ x_1 a_{m1} + \cdots + x_n a_{mn} \end{pmatrix} = \begin{pmatrix} \vec{A}_1 \cdot \mathbf{x} \\ \vdots \\ \vec{A}_m \cdot \mathbf{x} \end{pmatrix} \qquad \blacksquare$$

Remark 1.11 What's interesting about this definition is that if we take $\mathbf{x} \in \mathbb{R}^n$ and $A \in M_{m,n}(\mathbb{R})$, then $A\mathbf{x} \in \mathbb{R}^m$ is *uniquely* defined. This shows that *matrices in $M_{m,n}(\mathbb{R})$ may be thought of as functions from $\mathbb{R}^n$ to $\mathbb{R}^m$*,

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

[1] I didn't want to say the *row-size* of $\mathbf{x}$, because we actually need $\mathbf{x}$ to be in column form in the product itself—I just didn't want to have awkward text, with a giant column vector smack in the middle of the paragraph. Math, like fashion and art, is concerned with aesthetics—elegance is a goal!—even in spite of its practical uses. Nevertheless, it is important to know the difference between an *aesthetic choice*, such as laying $\mathbf{x}$ on its side *in the text*, and a *rigorously defined* form, such as the column vector form for $\mathbf{x}$ *in the matrix product* itself. This latter is a *definition*, and so includes the form in its presentation. We define things for ourselves and each other, after all, and we need to be clear, not only about the *words* which define things, but also the *form of their presentation* in the definition. Math is a *formal language*, no?

Very interesting! So a matrix, which is an array, may be thought of as a function. Well:

> *What kind* of function does a matrix define? This is a *typology* question, or a *definitional* question. It asks for a list of *properties* on which to pin a concept—if we get this right we'll get a bonafide *definition* of this concept, which is progress! OK, consider our situation, it's actually quite simple: Immediately from the definitions above, we can see that any matrix $A \in M_{m,n}(\mathbb{R})$ satisfies
>
> (1) $A(c\mathbf{x}) = cA\mathbf{x}$
>
> (2) $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y}$
>
> for all $c \in \mathbb{R}$ and all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (see proof below). Happily—or, more accurately, after enormous efforts in the 19th century, mathematicians began to agree that—*just these two properties suffice to define* a new class of functions, the ***linear functions*** (aka ***linear transformations***),
>
> $$\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$
>
> This is a subset of the set of *all possible* functions from $\mathbb{R}^n$ to $\mathbb{R}^m$,
>
> $$\mathcal{F}(\mathbb{R}^n, \mathbb{R}^m)$$
>
> From the category theory point of view, $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ *is the set to be studied* in linear algebra, because it contains all and only the *archetypal* or *characteristic* types of functions between vector spaces, algebraically speaking. In the *vector space* category, the objects are vector spaces (the $\mathbb{R}^n$s), but the *morphisms* or *functions*, the heart of the matter in this category, are the linear transformations. These we will study in depth in this class.

'What is there to study about $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$?,' you might ask. Well, now we have two sets,

$$M_{m,n}(\mathbb{R}) \qquad \text{and} \qquad \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$

What's the connection? And is it a connection we can exploit? The *connection* is of a certain mathematical type, namely that *between an abstract concept and its concrete realization*, for an abstract linear transformation needs to be *realized* in some sense to become definitely associated with a concrete matrix. *What is the nature of this realization?* Answer: ***coordinates***. Coordinates are the medium through which the abstract becomes concrete in the vector space category. ■

Remark 1.12 Look into the future: our desire for completeness and comprehensiveness doesn't stop at inventing linear transformations. We will need to add an array of ancillary concepts. For example, looking at the right hand side of 'def 1' of matrix multiplication in Definition 1.13, we can coin a word for the phenomenon: call the RHS of that equation, $\sum_{j=1}^{n} x_i \mathbf{a}_i$, a *linear combinations of the columns of A*, and call the set of all possible linear combinations the columns' *span*. With this technical vocabulary at hand, we can start comparing different things, for example we note that the *range of the function* A actually equals the *span* of the columns of A. This is the sort of conceptual criss-crossing that one hopes is cross-fertilizing—it is! ■

Definition 1.13 We can also **multiply two matrices** of the appropriate size. If A is an $m \times n$ matrix and B is an $n \times p$ matrix (m, n and p may be different here, but *note that n appears in both the dimensions of A and B, in specific locations!*), then the product of A and B, written AB, is an $m \times p$ matrix whose entries c_{ij} are the dot products of the ith row of A and the jth column of B:

$$AB = \begin{pmatrix} -\vec{A}_1- \\ -\vec{A}_2- \\ \vdots \\ -\vec{A}_m- \end{pmatrix} \begin{pmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \\ | & | & & | \end{pmatrix} \stackrel{\text{def 1}}{=} \begin{pmatrix} \vec{A}_1 \cdot \mathbf{b}_1 & \vec{A}_1 \cdot \mathbf{b}_2 & \cdots & \vec{A}_1 \cdot \mathbf{b}_p \\ \vec{A}_2 \cdot \mathbf{b}_1 & \vec{A}_2 \cdot \mathbf{b}_2 & \cdots & \vec{A}_2 \cdot \mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \vec{A}_m \cdot \mathbf{b}_1 & \vec{A}_m \cdot \mathbf{b}_2 & \cdots & \vec{A}_m \cdot \mathbf{b}_p \end{pmatrix}$$

What's nice about this definition is that it tells you what the ijth element of the product matrix AB is: let us write $(AB)_{ij}$ for the ijth term, then

$$(AB)_{ij} = \vec{A}_i \cdot \mathbf{b}_j$$

For example, the element in the 2,3 position (2nd row, 3rd column) is $\vec{A}_2 \cdot \mathbf{b}_3$.

Alternatively, we could use the definition of a matrix product with a column, as follows:

$$AB \stackrel{\text{def 2}}{=} \begin{pmatrix} | & | & & | \\ A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \\ | & | & & | \end{pmatrix}$$

where

$$A\mathbf{b}_j = b_{1j}\mathbf{a}_1 + b_{2j}\mathbf{a}_2 + \cdots + b_{nj}\mathbf{a}_n = \begin{pmatrix} \vec{A}_1 \cdot \mathbf{b}_j \\ \vdots \\ \vec{A}_m \cdot \mathbf{b}_j \end{pmatrix}$$

The jth column $A\mathbf{b}_j$ of AB has as it's ith entry the element $\vec{A}_i \cdot \mathbf{b}_j$, which is what our first definition asserted. That's why the two definitions agree. ■

Example 1.14 Let us see how this works when we multiply a 3×2 matrix with a 2×5. We should, of course, get a 3×5 matrix. Let

$$A = \begin{pmatrix} 1 & -1 \\ 0 & \sqrt{2} \\ 2 & -4 \end{pmatrix}, \qquad B = \begin{pmatrix} \sqrt{7} & 1 & 2 & 1 & 0 \\ 2 & -4 & 1 & 0 & 3 \end{pmatrix}$$

In terms of rows and columns, respectively,

$$A = \begin{pmatrix} -\vec{A}_1- \\ -\vec{A}_2- \\ -\vec{A}_3- \end{pmatrix} = \begin{pmatrix} \langle 1, -1 \rangle \\ \langle 0, \sqrt{2} \rangle \\ \langle 2, -4 \rangle \end{pmatrix}, \qquad B = \begin{pmatrix} | & | & | & | & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 & \mathbf{b}_5 \\ | & | & | & | & | \end{pmatrix}$$

$$= \left(\begin{pmatrix} \sqrt{7} \\ 2 \end{pmatrix} \begin{pmatrix} 1 \\ -4 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right)$$

which we now plug in:

$$\begin{aligned} AB &= \begin{pmatrix} -\vec{A}_1- \\ -\vec{A}_2- \\ -\vec{A}_3- \end{pmatrix} \begin{pmatrix} | & | & | & | & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 & \mathbf{b}_5 \\ | & | & | & | & | \end{pmatrix} \\ &= \begin{pmatrix} \vec{A}_1 \cdot \mathbf{b}_1 & \vec{A}_1 \cdot \mathbf{b}_2 & \vec{A}_1 \cdot \mathbf{b}_3 & \vec{A}_1 \cdot \mathbf{b}_4 & \vec{A}_1 \cdot \mathbf{b}_5 \\ \vec{A}_2 \cdot \mathbf{b}_1 & \vec{A}_2 \cdot \mathbf{b}_2 & \vec{A}_2 \cdot \mathbf{b}_3 & \vec{A}_2 \cdot \mathbf{b}_4 & \vec{A}_2 \cdot \mathbf{b}_5 \\ \vec{A}_3 \cdot \mathbf{b}_1 & \vec{A}_3 \cdot \mathbf{b}_2 & \vec{A}_3 \cdot \mathbf{b}_3 & \vec{A}_3 \cdot \mathbf{b}_4 & \vec{A}_3 \cdot \mathbf{b}_5 \end{pmatrix} \\ &= \begin{pmatrix} \langle 1,-1 \rangle \cdot \begin{pmatrix} \sqrt{7} \\ 2 \end{pmatrix} & \langle 1,-1 \rangle \cdot \begin{pmatrix} 1 \\ -4 \end{pmatrix} & \langle 1,-1 \rangle \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} & \langle 1,-1 \rangle \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \langle 1,-1 \rangle \cdot \begin{pmatrix} 0 \\ 3 \end{pmatrix} \\ \langle 0,\sqrt{2} \rangle \cdot \begin{pmatrix} \sqrt{7} \\ 2 \end{pmatrix} & \langle 0,\sqrt{2} \rangle \cdot \begin{pmatrix} 1 \\ -4 \end{pmatrix} & \langle 0,\sqrt{2} \rangle \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} & \langle 0,\sqrt{2} \rangle \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \langle 0,\sqrt{2} \rangle \cdot \begin{pmatrix} 0 \\ 3 \end{pmatrix} \\ \langle 2,-4 \rangle \cdot \begin{pmatrix} \sqrt{7} \\ 2 \end{pmatrix} & \langle 2,-4 \rangle \cdot \begin{pmatrix} 1 \\ -4 \end{pmatrix} & \langle 2,-4 \rangle \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} & \langle 2,-4 \rangle \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \langle 2,-4 \rangle \cdot \begin{pmatrix} 0 \\ 3 \end{pmatrix} \end{pmatrix} \\ &= \boxed{\begin{pmatrix} \sqrt{7}-2 & 5 & 1 & 1 & -3 \\ 2\sqrt{2} & -4\sqrt{2} & \sqrt{2} & 0 & 3\sqrt{2} \\ 2\sqrt{7}-8 & 18 & 0 & 2 & -12 \end{pmatrix}} \end{aligned}$$

■

Notation 1.15 *We have now finished defining matrices, in a certain sense. The set $M_{m,n}(\mathbb{R})$ of all $m \times n$ real matrices is the set of all $m \times n$ arrays, sure, but we also enriched it with extra structure, the algebraic structure of addition and scalar multiplication. This is part of the definition of $M_{m,n}(\mathbb{R})$, and we have special notation to indicate that $M_{m,n}(\mathbb{R})$ has more than just 'things,' matrices, in it: denoting addition by $+$ and scalar multiplication by $\cdot_s$, we can indicate this extra structure with a list:*

$$\boxed{\left(M_{m,n}(\mathbb{R}),\ +,\ \cdot_s\right)}$$

The list gives the set plus the two extra operations associated with this set (by definition). ■

Remark 1.16 We also added matrix multiplication to our list of propetries, but this doesn't make sense for a single set $M_{m,n}(\mathbb{R})$, unless $m = n$. Rather, *matrix multiplication is a two-variable map*:

$$\mu : M_{m,n}(\mathbb{R}) \times M_{n,p}(\mathbb{R}) \to M_{m,p}(\mathbb{R})$$
$$\mu(A, B) := AB$$

How should we interpret this? How does this fit into our vector space category? *Answer:* **function composition**. If $A \in M_{m,n}(\mathbb{R})$, then as a function

$$A : \mathbb{R}^n \to \mathbb{R}^m, \qquad \mathbf{y} \mapsto A\mathbf{y}$$

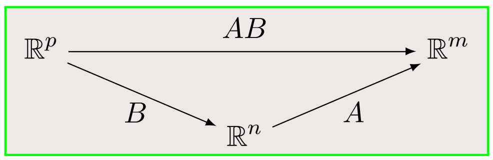
while if $B \in M_{n,p}(\mathbb{R})$, then as a function

$$B : \mathbb{R}^p \to \mathbb{R}^n, \qquad \mathbf{x} \mapsto B\mathbf{x}$$

so the composite must be a function

$$AB : \mathbb{R}^p \to \mathbb{R}^m, \qquad \begin{array}{l} \mathbf{x} \mapsto \mathbf{y} = B\mathbf{x} \ \text{ followed by} \\ \mathbf{y} \mapsto \mathbf{z} = A\mathbf{y} = A(B\mathbf{x}) = (AB)\mathbf{x} \end{array}$$

This is neatly summarized in the diagram,



This all becomes interesting when $m = n = p$, i.e. when A and B are square. For then matrix multiplication, $\cdot$, gives the set $M_n(\mathbb{R})$ an additional algebraic operation, and hence a new property:

$$\boxed{\big(M_n(\mathbb{R}),\ +,\ \cdot_s,\ \cdot\big)}$$

giving square matrices a similar structure to the integers,

$$n + m \quad \longleftrightarrow \quad A + B, \quad an \quad \longleftrightarrow \quad aA, \quad nm \quad \longleftrightarrow \quad AB$$

along with the corresponding properties, like associativity and distributivity. We prove all of this below. ■

One thing in the remark above bothers me, however. Maybe it's not so obvious that $A(B\mathbf{x}) = (AB)\mathbf{x}$. As it turns out, we need to prove this, and it takes a little work—but not too much. See Proposition 2.5 below.

2 Immediate Implications of the Algebraic Properties of Matrices

Theorem 2.1 (Matrix Addition) *For all matrices $A, B, C \in M_{m,n}(\mathbb{R})$, we have*

(1) $A + B = B + A$ *(commutativity of addition)*

(2) $(A + B) + C = A + (B + C)$ *(associativity of addition)*

(3) *There is a* **zero matrix** *in* $M_{m,n}(\mathbb{R})$

$$\boxed{O = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}}$$

which acts as, and is characterized by being, the ***additive identity***

$$\boxed{O + A = A + O = A}$$

for all $A \in M_{m,n}(\mathbb{R}^n)$.

(4) *Every* $A \in M_{m,n}(\mathbb{R})$ *has a* **negative** *or* **additive inverse**

$$\boxed{-A = \begin{pmatrix} -a_{11} & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{m1} & \cdots & -a_{mn} \end{pmatrix}}$$

Proof: Exercise. It's totally straightforward, model it on the analogous theorem for vector addition. Not surprisingly, the reason we're including it here is to later show that $M_{m,n}(\mathbb{R})$ *is a vector space!* ■

Theorem 2.2 (Matrix Scalar Multiplication) *For all $a, b \in \mathbb{R}$ and all $A, B \in M_{m,n}(\mathbb{R})$, we have*

(1) $a(bA) = (ab)A$ *(associativity of scalar mult.)*

(2) $(a + b)A = aA + bA$ *(distributivity over $\mathbb{R}$-addition)*

(3) $a(A + B) = aA + aB$ *(distributivity over $M_{m,n}(\mathbb{R})$-addition)*

(4) $1A = A$. ■

Proof: This is an easy exercise in chasing definitions, so we'll prove (1) and leave the rest for you. Since scalar multiplication is elementwise, it suffices to check each side's ijth element, $1 \leq i \leq m$, $1 \leq j \leq n$: the ijth element on the LHS is $a(ba_{ij})$, while on the RHS it is $(ab)a_{ij}$. But these two are equal in $\mathbb{R}$ because $\mathbb{R}$ enjoys associativity. Applying this reasoning simultaneously to all slots we have the result. ■

Lemma 2.3 *For all $A \in M_{m,n}(\mathbb{R})$, all $\mathbf{x} \in \mathbb{R}^n$ and all $c \in \mathbb{R}$, we have*

$$c(A\mathbf{x}) = (cA)\mathbf{x} = A(c\mathbf{x})$$

Proof: We simply compute using our definitions of matrix-vector product and algebraic properties of vectors:

$$\begin{aligned}
\boxed{c(A\mathbf{x})} &= c\left(\begin{pmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) \\
&= c\big(x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n\big) && \textit{(Definition 1.10)} \\
&= c(x_1\mathbf{a}_1) + \cdots + c(x_n\mathbf{a}_n) && \textit{(distributivity)} \\
&= (cx_1)\mathbf{a}_1 + \cdots + (cx_n)\mathbf{a}_n && \textit{(associativity)} \\
&= \begin{pmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{pmatrix}\begin{pmatrix} cx_1 \\ \vdots \\ cx_n \end{pmatrix} && \textit{(Definition 1.10)} \\
&= \boxed{A(c\mathbf{x})}
\end{aligned}$$

or, picking up the thread at the 4th row and heading in a different direction

$$\begin{aligned}
&\vdots \\
&= (cx_1)\mathbf{a}_1 + \cdots + (cx_n)\mathbf{a}_n \\
&= (x_1c)\mathbf{a}_1 + \cdots + (x_nc)\mathbf{a}_n && \textit{(commutativity in } \mathbb{R}\textit{)} \\
&= x_1(c\mathbf{a}_1) + \cdots + x_n(c\mathbf{a}_n) && \textit{(associativity)} \\
&= \begin{pmatrix} | & & | \\ c\mathbf{a}_1 & \cdots & c\mathbf{a}_n \\ | & & | \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} && \textit{(Definition 1.10)} \\
&= \boxed{(cA)\mathbf{x}}
\end{aligned}$$

This completes the proof. ■

We single out only part of the previous Lemma, and insert it into property (1) below, not to be irritating, but to emphasize the *pair* of properties below as a unit. They will become the *defining properties* of a linear transformation (Definition 5.1 below). There is a very general way to prove the second property (2), for functions in general. Instead, we'll prove it directly, as an exercise in matrix calculations.

Theorem 2.4 (Matrices as Linear Transformations) *Let* $A \in M_{m,n}(\mathbb{R})$*, and think of it as a function* $\mathbb{R}^n \to \mathbb{R}^m$*. Then, for all* $c \in \mathbb{R}$ *and all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*, scalar multiplication* $\cdot_s$ *and addition* $+$ *satisfy*

(1) $A(c\mathbf{x}) = c(A\mathbf{x})$

(2) $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y}$

Proof: We proved (1) in the Lemma, so it remains to prove (2): Let $A \in M_{m,n}(\mathbb{R})$ and choose arbitrary vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then,

$$\begin{aligned}
A(\mathbf{x}+\mathbf{y}) &= \begin{pmatrix} \mid & & \mid \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \mid & & \mid \end{pmatrix} \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \\
&= (x_1+y_1)\mathbf{a}_1 + \cdots + (x_n+y_n)\mathbf{a}_n \qquad \textit{(Definition 1.10)} \\
&= (x_1\mathbf{a}_1 + y_1\mathbf{a}_1) + \cdots + (x_n\mathbf{a}_n + y_n\mathbf{a}_n) \qquad \textit{(distributivity)} \\
&= \big(x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n\big) + \big(y_1\mathbf{a}_1 + \cdots + y_n\mathbf{a}_n\big) \qquad \textit{(associativity)} \\
&= \begin{pmatrix} \mid & & \mid \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \mid & & \mid \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \mid & & \mid \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \mid & & \mid \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad \textit{(Definition 1.10)} \\
&= A\mathbf{x} + A\mathbf{y}
\end{aligned}$$

which finishes the proof. ■

We finally return to our qualms about Remark 1.16 (see the comments following that remark), and prove that nagging claim. It's important, otherwise that whole remark fails.

Proposition 2.5 *For all* $A \in M_{m,n}(\mathbb{R})$*, all* $B \in M_{n,p}(\mathbb{R})$ *and all* $\mathbf{x} \in \mathbb{R}^p$*, we have*

$$A(B\mathbf{x}) = (AB)\mathbf{x} \tag{2.1}$$

Proof: We simply compute from left to right, using Definition 1.13, version 2 for multiplication, and otherwise only the work we've already developed, like this:

$$\begin{aligned}
(AB)\mathbf{x} &= \begin{pmatrix} \mid & & \mid \\ A\mathbf{b}_1 & \cdots & A\mathbf{b}_p \\ \mid & & \mid \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \qquad \textit{(Definition 1.13, def 2)} \\
&= x_1(A\mathbf{b}_1) + \cdots + x_p(A\mathbf{b}_p) \qquad \textit{(Definition 1.10)} \\
&= A(x_1\mathbf{b}_1) + \cdots + A(x_p\mathbf{b}_p) \qquad \textit{(Lemma 2.3)} \\
&= A\big(x_1\mathbf{b}_1 + \cdots + x_p\mathbf{b}_p\big) \qquad \textit{(Theorem 2.4)} \\
&= A\left(\begin{pmatrix} \mid & & \mid \\ \mathbf{b}_1 & \cdots & \mathbf{b}_p \\ \mid & & \mid \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}\right) \qquad \textit{(Definition 1.10)} \\
&= A(B\mathbf{x})
\end{aligned}$$

thus completing the proof. ■

Theorem 2.6 (Matrix Multiplication) *For any matrices A, B and C whose dimensions are compatible, and for every $c \in \mathbb{R}$ we have*

(0) $AB \neq BA$ *in general, even when defined.* (non-commutativity of addition)

(1) $A(BC) = (AB)C$ *(associativity of multiplication)*

(2) $A(B + C) = AB + AC$ *(distributivity from the left)*

(3) $(A + B)C = AC + BC$ *(distributivity from the right)*

(4) $c(AB) = (cA)B = A(cB)$

(5) *There is an **identity** matrix,*

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

*which acts as, and is characterized by being, the **multiplicative identity**,*

$$AI_n = A = I_m A$$

Proof: (0) To see this, consider a random example: $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} -2 & 1 \\ 3 & -4 \end{pmatrix}$. Then,

$$AB = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 3 & -4 \end{pmatrix} = \begin{pmatrix} 4 & -7 \\ 6 & -13 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ -9 & -10 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 3 & -4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = BA$$

(1) In this case, $A \in M_{m,n}(\mathbb{R})$, $B \in M_{n,p}(\mathbb{R})$ and $C \in M_{p,r}(\mathbb{R})$. We use Theorem 2.4 (2), Proposition 2.5 and Definition 1.13 (def 2):

$$\begin{aligned} A(BC) = A\big(B\mathbf{c}_1 \quad \cdots \quad B\mathbf{c}_r\big) &= \big(A(B\mathbf{c}_1) \quad \cdots \quad A(B\mathbf{c}_r)\big) \\ &= \big((AB)\mathbf{c}_1 \quad \cdots \quad (AB)\mathbf{c}_r\big) = (AB)C \end{aligned}$$

(2) In this case, $A \in M_{m,n}(\mathbb{R})$, $B, C \in M_{n,p}(\mathbb{R})$. We use Theorem 2.4 (2) and Definition 1.13 (def 2):

$$\begin{aligned} A(B + C) &= A\big((\mathbf{b}_1 + \mathbf{c}_1) \quad \cdots \quad (\mathbf{b}_p + \mathbf{c}_p)\big) \\ &= \big(A(\mathbf{b}_1 + \mathbf{c}_1) \quad \cdots \quad A(\mathbf{b}_p + \mathbf{c}_p)\big) = \big(A\mathbf{b}_1 + A\mathbf{c}_1 \quad \cdots \quad A\mathbf{b}_p + A\mathbf{c}_p\big) \\ &= \big(A\mathbf{b}_1 \quad \cdots \quad A\mathbf{b}_p\big) + \big(A\mathbf{c}_p \quad \cdots \quad A\mathbf{c}_p\big) = AB + AC \end{aligned}$$

(3) Similar to (2): Exercise!

(4) We use Lemma 2.3 and Definition 1.13 (def 2):

$$\begin{aligned} c(AB) = c\big(A\mathbf{b}_1 \quad \cdots \quad A\mathbf{b}_p\big) &= \big(c(A\mathbf{b}_1) \quad \cdots \quad c(A\mathbf{b}_p)\big) \\ &= \big((cA)\mathbf{b}_1 \quad \cdots \quad (cA)\mathbf{b}_p\big) = (cA)B \end{aligned}$$

and similarly

$$\begin{aligned} c(AB) = c\big(A\mathbf{b}_1 \quad \cdots \quad A\mathbf{b}_p\big) &= \big(c(A\mathbf{b}_1) \quad \cdots \quad c(A\mathbf{b}_p)\big) \\ &= \big(A(c\mathbf{b}_1) \quad \cdots \quad A(c\mathbf{b}_p)\big) = A\big((c\mathbf{b}_1) \quad \cdots \quad (c\mathbf{b}_p)\big) = A(cB) \end{aligned}$$

(5) The columns of I_n are $\mathbf{e}_1 = (1, 0, \ldots, 0)^T, \ldots, \mathbf{e}_n = (0, \ldots, 0, 1)^T$, so since

$$A\mathbf{e}_j = \begin{pmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = 0\mathbf{a}_1 + \cdots + 1\mathbf{a}_j + \cdots + 0\mathbf{a}_n = \mathbf{a}_j$$

we have

$$AI_n = \big(A\mathbf{e}_1 \quad \cdots \quad A\mathbf{e}_n\big) = \big(\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n\big) = A$$

A similar argument, necessarily modified, proves $I_m A = A$. Do it as an exercise! ■

3 Related Dichotomies: Null Space/Range, One-to-One/ Onto, and also Column Space

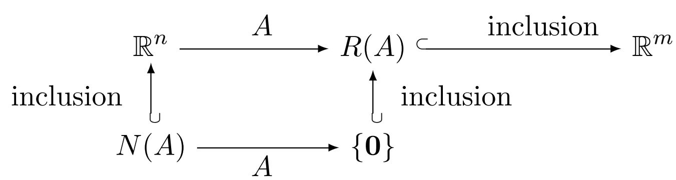
Definition 3.1 Let $A \in M_{m,n}(\mathbb{R})$, but consider it as a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Associated to *any* function is its ***range*** or ***image***, the *set of achieved* $\mathbf{y}$*-values*, and so here:

$$\boxed{R(A) \;\text{ or }\; A(\mathbb{R}^n) \;\text{ or }\; \operatorname{im} A = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = A\mathbf{x} \;\text{ for some } \mathbf{x} \in \mathbb{R}^n\}}$$

Because matrix functions are also *linear* (Theorem 2.4), there is a *second* natural space associated to A, the ***null space*** or ***kernel***, which is the set of domain-vectors $\mathbf{x}$ sent to $\mathbf{0}$:

$$\boxed{N(A) \;\; \textit{or} \;\; \ker A = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}}$$

Note: *$N(A)$ is a subset of $\mathbb{R}^n$ while $R(A)$ is a subset of $\mathbb{R}^m$,*



But a happy thing happens on account of the fact that matrices are *linear* functions: the subsets $N(A) \subseteq \mathbb{R}^n$ and $R(A) \subseteq \mathbb{R}^m$ are actually **sub*spaces***: ■

Theorem 3.2 *Let $A \in M_{m,n}(\mathbb{R})$. Then, $N(A)$ is a sub*space *of $\mathbb{R}^n$ and $R(A)$ is a sub*space *of $\mathbb{R}^m$.*

Proof: (1) We verify that if $\mathbf{x}_1, \mathbf{x}_2 \in N(A)$ and $c \in \mathbb{R}$, then $\mathbf{x}_1 + \mathbf{x}_2 \in N(A)$ and $c\mathbf{x}_1 \in N(A)$. But this will follow directly from the ability to distribute, Theorem 2.4:

$$\begin{aligned} \mathbf{x}_1, \mathbf{x}_2 \in N(A) \quad &\Longrightarrow \quad A\mathbf{x}_1 = A\mathbf{x}_2 = \mathbf{0} \\ &\Longrightarrow \quad A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0} \\ &\Longrightarrow \quad \mathbf{x}_1 + \mathbf{x}_2 \in N(A) \end{aligned}$$

Similarly, by that same theorem, we can pull the scalar out and use the assumption:

$$\begin{aligned} c \in \mathbb{R} \;\text{ and }\; \mathbf{x}_1 \in N(A) \quad &\Longrightarrow \quad A(c\mathbf{x}_1) = c(A\mathbf{x}_1) = c\mathbf{0} = \mathbf{0} \\ &\Longrightarrow \quad c\mathbf{x}_1 \in N(A) \end{aligned}$$

This completes the proof that $N(A)$ is a subspace of $\mathbb{R}^n$.

(2) Let's look at the range, now. Let $\mathbf{y}_1, \mathbf{y}_2 \in R(A)$ and $c \in \mathbb{R}$. We need to show that $\mathbf{y}_1 + \mathbf{y}_2 \in R(A)$ and $c\mathbf{y}_1 \in R(A)$.

$$\begin{aligned} \mathbf{y}_1, \mathbf{y}_2 \in R(A) \quad &\Longrightarrow \quad \mathbf{y}_1 = A\mathbf{x}_1 \;\text{ and }\; \mathbf{y}_2 = A\mathbf{x}_2 \;\text{ for some }\; \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n \\ &\Longrightarrow \quad \mathbf{y}_1 + \mathbf{y}_2 = A\mathbf{x}_1 + A\mathbf{x}_2 = A(\mathbf{x}_1 + \mathbf{x}_2) \\ &\Longrightarrow \quad \mathbf{y}_1 + \mathbf{y}_2 \in R(A), \;\; \text{since } \mathbf{x}_1 + \mathbf{x}_2 \in \mathbb{R}^n \end{aligned}$$

Similarly,

$$\begin{aligned} c \in \mathbb{R} \ \text{ and } \ \mathbf{y}_1 \in R(A) \quad &\Longrightarrow \quad c\mathbf{y}_1 = c(A\mathbf{x}_1) = A(c\mathbf{x}_1) \\ &\Longrightarrow \quad c\mathbf{y}_1 \in R(A) \end{aligned}$$

since $c\mathbf{x}_1 \in \mathbb{R}^n$. ■

Remark 3.3 The further analysis of $N(A)$ and $R(A)$, including their coordinatization, will have to wait until we explain how row-reduction actually works. That's why we had to first develop the algebraic mechanics of matrices and vectors, because row-reduction is otherwise a weird mystery with a coincidental happy ending. Matrices *clarify the underlying structure* of the method—namely, the *conceptual structure*. Hopefully, if we do this right, this conceptual structure will give *proof* and *explanation* of the correctness of the row-reduction method. ■

Remark 3.4 We can at least say a couple of things *right now* about the range, as straight consequences of the definitions. What we can do is build a bridge between the concept of the ***column space*** of $A = \begin{pmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{pmatrix} \in M_{m,n}(\mathbb{R})$, which is the *span of the columns of* A,

$$\boxed{\mathrm{Col}(A) := \mathrm{span}\begin{pmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{pmatrix}}$$

and two other *seemingly* unrelated questions that will, in fact, turn out to be of interest when we *do* go into row-reduction a little more rigorously: whether or not a system $A\mathbf{x} = \mathbf{b}$ is ***consistent*** (meaning *has* a solution) and whether or not the vector $\mathbf{b}$ has anything to do with it. The bridge will be constructed out of the *function*-aspect of A, its ***range***, and it goes like this: if the range $R(A)$ is all of $\mathbb{R}^m$ (i.e. A is onto), the system will be consistent no matter what the $\mathbf{b}$, but if $R(A)$ is not all of $\mathbb{R}^m$, then consistency is precisely the question of whether $\mathbf{b} \in R(A)$ or $\notin R(A)$. ■

We begin with the simple observation that the range and the column space are the same thing.

Proposition 3.5 *Let* $A = \begin{pmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{pmatrix} \in M_{m,n}(\mathbb{R})$*, and think of it as a function* $\mathbb{R}^n \to \mathbb{R}^m$*. Then, its range* $R(A)$ *may be characterized as the span of its columns,*

$$\boxed{R(A) = \mathrm{span}\big(\mathbf{a}_1, \ldots, \mathbf{a}_n\big) = \mathrm{Col}(A)}$$

Proof: By definition, the range of A is the set of *achieved* $\mathbf{y}$-values,

$$R(A) = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = A\mathbf{x} \ \text{ for some } \ \mathbf{x} \in \mathbb{R}^n\}$$

while, by definition, the column space of A is the span, i.e. *set of all possible linear combinations*, of the columns,

$$\mathrm{Col}(A) = \mathrm{span}\big(\mathbf{a}_1, \ldots, \mathbf{a}_n\big) \stackrel{\text{def}}{=} \left\{ \sum_{i=1}^n c_i \mathbf{a}_i \ \middle| \ c_1, \ldots, c_n \in \mathbb{R} \right\}$$

Our definition of matrix products with vectors, Definition 1.10, is the link, for by that definition the term $A\mathbf{x}$ in the defining property of the range, $\mathbf{y} = A\mathbf{x}$, is a linear combination of the columns: thus,

$$\underbrace{\mathbf{y} = A\mathbf{x}}_{\in R(A)} = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \stackrel{\text{def}}{=} \underbrace{x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n}_{\in \operatorname{span}(\mathbf{a}_1,\ldots,\mathbf{a}_n)}$$

and the RHS of this is nothing if not a linear combination of the columns of A, i.e. *belonging to the column space of* A. So we conclude that all of the $\mathbf{y}$-values in the $R(A)$ are also in $\operatorname{Col}(A)$, and vice-versa. ■

Remark 3.6 This can have two interpretations: one, that the *span of its columns characterizes the range of* A, and two, that *the range of* A *characterizes the span of its columns*. Which is it? I like to think it is the latter, though, again, the two are logically equivalent:

> By understanding A as a function, which provides one with the notion of *range*, we can *rephrase* the *consistency question*—whether or not $A\mathbf{x} = \mathbf{b}$ has a solution—which is conceptually rather poverty-stricken, into a conceptually rich question of analyzing a function's range, along with a way to do it, namely by looking at the span of the columns and using vector space concepts. We repeat here, however, that the *full* analysis of $R(A)$ and its role in finding *all solutions* of $A\mathbf{x} = \mathbf{b}$ will have to wait until we add more conceptual structure to our picture.

With an eye towards this later development, let us now go ahead and give some preliminary results.

■

Proposition 3.7 *Let* $A \in M_{m,n}(\mathbb{R})$*, and let* $\mathbf{b} \in \mathbb{R}^m$*. Then the following statements are logically equivalent:*

(1) $\mathbf{b} \in R(A)$

(2) $\mathbf{b} \in \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \operatorname{Col}(A)$

(3) $A\mathbf{x} = \mathbf{b}$ *is consistent/has a solution.*

Proof: In the light of Proposition 3.5, the equivalence of (1) and (2) is clear, since $R(A) = \operatorname{Col}(A)$: if $\mathbf{b}$ is in one, it is in the other and vice-versa. But the equivalence of (1) and (3) is the definition of $R(A)$: if $\mathbf{b} \in R(A)$, then $\mathbf{b} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, in which case there *is* a solution, namely $\mathbf{x}$, for the system $A\mathbf{x} = \mathbf{b}$. Conversely, if $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x} \in \mathbb{R}^n$, then $\mathbf{b}$ is the $\mathbf{y}$-value of A, namely the $\mathbf{y}$-value gotten by 'plugging in' $\mathbf{x}$ into A. ■

Corollary 3.8 (Characterizations of Onto) *Let $A \in M_{m,n}(\mathbb{R})$. Then the following statements are logically equivalent, and therefore serve to characterize an onto matrix function:*

(1) *A is onto, i.e. $R(A) = \mathbb{R}^m$.*

(2) $\text{Col}(A) = \text{span}\big(\mathbf{a}_1, \ldots, \mathbf{a}_n\big) = \mathbb{R}^m$

(3) *$A\mathbf{x} = \mathbf{b}$ is consistent/has a solution* for all *$\mathbf{b} \in \mathbb{R}^m$.*

Proof: This corollary just makes clear the *consequence* of the above Proposition 3.7: Now that we've clarified the role of $\mathbf{b}$ in relation to $R(A)$ and consistency, the only question left is *whether or not* $\mathbf{b}$ *is in fact in the range of* A: does $\mathbf{b}$ lie in $R(A)$ or is it merely in $\mathbb{R}^m$ but not in $R(A)$, symbolically

$$\boxed{\mathbf{b} \in R(A) \quad \text{or} \quad \mathbf{b} \in \mathbb{R}^m \backslash R(A)\,?}$$

But this is the same question as

$$\boxed{\textit{Is } A \textit{ onto or not?}}$$

For, in order that A be onto, we recall, we must have that *all possible* $\mathbf{y}$*-values, all vectors in* $\mathbb{R}^n$*, are achieved, i.e.* $R(A) = \mathbb{R}^m$. ■

We can also say something *right now about injectivity, or whether A is one-to-one*, from purely basic facts about the columns and the null space of A.

Proposition 3.9 (Characterizations of One-to-One) *Let $A \in M_{m,n}(\mathbb{R})$. Then the following statements are logically equivalent, and therefore serve to characterize a one-to-one matrix function:*

(1) *A is one-to-one, i.e. $A\mathbf{x}_1 = A\mathbf{x}_2 \implies \mathbf{x}_1 = \mathbf{x}_2$.*

(2) *$N(A) = \{\mathbf{0}\}$.*

(3) *The columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of A are linearly independent in $\mathbb{R}^m$.*

Proof: Let us show that (1) implies (2): Take a one-to-one matrix A in hand, and pick a random $\mathbf{x} \in N(A)$, i.e. an $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{0}$. Since certainly $A\mathbf{0} = \mathbf{0}$, we make the innocent-looking observation that $A\mathbf{x} = \mathbf{0} = A\mathbf{0}$ seconds before we cash in our one-to-one assumption for our desired conclusion: $\mathbf{x} = \mathbf{0}$. We've just shown that if $\mathbf{x} \in N(A)$, then $\mathbf{x} = \mathbf{0}$, which in plain English says that $N(A) = \{\mathbf{0}\}$.

Next, the converse, (2) implies (1): Take a matrix A with trivial null space, $N(A) = \{\mathbf{0}\}$ and see what happens when you try to set $A\mathbf{x}_1 = A\mathbf{x}_2$. You'll find Theorem 2.4 useful here, for distributing:

$$\begin{aligned} A\mathbf{x}_1 = A\mathbf{x}_2 \quad &\implies \quad A(\mathbf{x}_1 - \mathbf{x}_2) = A\mathbf{x}_1 - A\mathbf{x}_2 = \mathbf{0} \quad \textit{(subtract and factor } A\textit{)} \\ &\implies \quad \mathbf{x}_1 - \mathbf{x}_2 \in N(A) = \{\mathbf{0}\} \quad \textit{(last step shows this)} \\ &\implies \quad \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0} \quad \textit{(since } N(A) \textit{ contains only } \mathbf{0}\textit{)} \\ &\implies \quad \mathbf{x}_1 = \mathbf{x}_2 \end{aligned}$$

Next, let's show that (2) imples (3): If $N(A) = \{\mathbf{0}\}$, then

$$\begin{aligned} x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0} \quad &\Longrightarrow \quad A\mathbf{x} = \begin{pmatrix}\mathbf{a} & \cdots & \mathbf{a}\end{pmatrix}\begin{pmatrix}x_1 \\ \vdots \\ x_n\end{pmatrix} \overset{\text{Def 1.10}}{=} x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0} \\ &\Longrightarrow \quad \mathbf{x} \in N(A) = \{\mathbf{0}\} \\ &\Longrightarrow \quad \mathbf{x} = \mathbf{0} \\ &\Longrightarrow \quad x_1 = \cdots = x_n = 0 \end{aligned}$$

which shows that the columns of A are linearly indepedent.

The converse, (3) implies (2), goes like this: If the columns of A are linearly independent, then we reverse the above chain of reasoning:

$$\begin{aligned} \mathbf{x} \in N(A) \quad &\Longrightarrow \quad A\mathbf{x} = \mathbf{0} \\ &\Longrightarrow \quad x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \begin{pmatrix}\mathbf{a} & \cdots & \mathbf{a}\end{pmatrix}\begin{pmatrix}x_1 \\ \vdots \\ x_n\end{pmatrix} \underbrace{= A\mathbf{x} = \mathbf{0}}_{\text{(by last line)}} \\ &\Longrightarrow \quad x_1 = \cdots = x_n = 0 \quad \textit{(since the columns of } A \textit{ are lin. indep.)} \\ &\Longrightarrow \quad \mathbf{x} = \mathbf{0} \end{aligned}$$

which shows that $N(A) = \{\mathbf{0}\}$. ■

4 Both One-to-One and Onto: Invertible Matrices

4.1 General Properties of Invertible Matrices

What do the results of the last section show? Let's compare them: for any matrix $A \in M_{m,n}(\mathbb{R})$,

A is **one-to-one**	$\iff$	the columns of A are **linearly independent** in $\mathbb{R}^m$
	$\iff$	$N(A) = \{\mathbf{0}\} \subseteq \mathbb{R}^n$
		VS
A is **onto**	$\iff$	the columns of A **span** $\mathbb{R}^m$
	$\iff$	$R(A) = \mathbb{R}^m$

Remark 4.1 There are certain ***barriers*** to a matrix A being one-to-one and other barriers to it being onto. The *simplest* or *easiest to identify* of these are the *dimensions*, m and n:

Claim 1: *A cannot be one-to-one if* $m < n$.

Claim 2: *A cannot be onto if* $m > n$.

We have stated these claims in class and used them on homeworks, quizzes, and exams, but have only hinted at proofs, arguing vaguely about the number of pivot and non-pivot columns, or equivalently the number of independent vs. dependent variables. *These numbers can in fact be used to* confirm *or* deny *either of the two claims above.* But the *precise* reason for their truth must await the **Rank-Nullity Theorem** (in our book called the Rank Theorem), or else await a better explanation of row-reduction and pivots. We will do both of these.

Thus, let us remark for now that we know, though we have yet to prove, that the *only chance a matrix has of being invertible (both one-to-one and onto) is if* $m = n$, *i.e. if A is* ***square***. ■

With this in mind, the only logical conclusion to draw from our work above is:

$A \in M_n(\mathbb{R})$ is **invertible**	$\overset{\text{def}}{\iff}$	A is **both one-to-one and onto**
	$\iff$	the columns of A are both **linearly independent and span** $\mathbb{R}^n$ (Corollary 3.8 & Proposition 3.9)
	$\overset{\text{def}}{\iff}$	the columns of A form a **basis** $\alpha = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ for $\mathbb{R}^n$

Very interesting... Invertibility of a square matrix is *logically equivalent* to its columns forming a basis for $\mathbb{R}^n$. This must be the basis gotten from the *standard basis* $\sigma = \big(\mathbf{e}_1, \dots, \mathbf{e}_n\big)$ by multiplying it through by A (recalling, from the proof of (5) in Theorem 2.6, that $A\mathbf{e}_i = \mathbf{a}_i$):

$$\boxed{\alpha = \big(\mathbf{a}_1, \dots, \mathbf{a}_n\big) = \big(A\mathbf{e}_1, \dots, A\mathbf{e}_n\big) = A\sigma}$$

Thus, **to say α is a basis for $\mathbb{R}^n$ *is* to say that the matrix having its vectors for columns is invertible.** There is a connection between bases and invertible matrices that will have to be explored further.

Definition 4.2 The set of all *invertible* matrices A in $M_n(\mathbb{R})$ has a name, the (real) **general linear group**, and is denoted

$$\boxed{\mathrm{GL}(n, \mathbb{R}) := \{A \in M_n(\mathbb{R}) \mid A \ \text{ is invertible (both one-to-one and onto)}\}}$$ ■

Remark 4.3 The 'linear' part of the name is clear, since matrix multiplication is linear (Theorem 2.4). The 'general' part of the name refers to the 'all' in 'the set of all' part of the definition. So it is only the *'group'* part that needs explaining. ■

Definition 4.4 A **group** is a set G with an **identity element** $e \in G$ and two operations, **group multiplication**

$$\boxed{g, h \ \mapsto \ gh}$$

and **group inversion**,

$$\boxed{g \ \mapsto g^{-1}}$$

for all $g \in G$. That is, *every element in G is invertible*, and this can only be understood *with respect to the identity e*, as the following *defining* identity:

$$\boxed{g^{-1}g = gg^{-1} = e}$$

These operations are subject to certain familiar rules:[2]

(i) $\boxed{eg = ge = g}$ for all $g \in G$. (e ***is a multiplicative identity***)

(ii) $\boxed{g(hk) = (gh)k}$ for all $g, h, k \in G$ (***associativity***)

for all $g \in G$. As usual, to indicate the fact that the *group* G is a set with *extra structure*—the algebraic structure of multiplication $\boldsymbol{\cdot}$, inversion ${}^{-1}$, and an identity e—we make a list:

$$\boxed{(G, \ \boldsymbol{\cdot}, \ {}^{-1}, \ e)}$$ ■

[2]Familiar, that is, because we always proceed from the more familiar to the less familiar, and the more familiarity we're referring to here is the very lists of properties we wrote down earlier in Theorems 2.1 and 2.6.

Remark 4.5 We *do not* assume commutativity. That is exactly the thing that fails for matrices, anyway—coincidence? No, it's *because* commutativity fails for matrices that we don't include it. The abstract group is *inspired* by the general linear group of invertible $n \times n$ matrices, not the other way around. ■

Clearly $\mathrm{GL}(n, \mathbb{R})$ is a group, with identity e the identity matrix I_n, group multiplication given by matrix multiplication, and inversion given by the *function*-inverse. Associativity and the identity $IA = AI = A$ were proved in Theorem 2.6, parts (1) and (5) (applied to the case $m = n$ here). What else can we say? In the way of general properties (applicable to *all* groups, in fact), we can say this much:

Theorem 4.6 *Let $A, B \in \mathrm{GL}(n, \mathbb{R})$ be invertible square matrices, and let $a \neq 0$ be any nonzero real number. Then, A^{-1}, aA, AB, and $A^T \in \mathrm{GL}(n, \mathbb{R})$, and*

(1) $(A^{-1})^{-1} = A$.

(2) $(aA)^{-1} = a^{-1}A^{-1}$.

(3) $(AB)^{-1} = B^{-1}A^{-1}$.

(4) $(A^T)^{-1} = (A^{-1})^T$.

Proof: These statements are of such a general nature that we don't really need to compute anything to prove them. They're easier to see *directly from the defining properties of* $\mathrm{GL}(n, \mathbb{R})$ *and our earlier theorems.*

(1) To see that A^{-1} is invertible, for example, we just stare at the identity

$$AA^{-1} = A^{-1}A = I$$

and see within it the desired conclusion, as follows: One of the characterizations of invertibility, for *any* function $f : X \to Y$, is the *existence* of another function $g : Y \to X$ satisfying

$$f \circ g = g \circ f = \mathrm{id}$$

i.e. $f(g(y)) = y$ and $g(f(x)) = x$. This just says f is the inverse of g iff g is the inverse of f:

$$g = f^{-1} \text{ iff } f = g^{-1}$$

Well, the above equality $AA^{-1} = A^{-1}A = I$ means (on account of Proposition 2.5) $A(A^{-1}\mathbf{y}) = \mathbf{y}$ and $A^{-1}(A\mathbf{x}) = \mathbf{x}$, so we conclude that A is A^{-1}'s inverse as much as A^{-1} is A's:

$$(A^{-1})^{-1} = A$$

(2) Similarly, from Theorem 2.6, part (4), we have

$$(a^{-1}A)(aA) = (a^{-1}a)(A^{-1}A) = I = (aa^{-1})(AA^{-1}) = (aA)(a^{-1}A^{-1})$$

and we conclude that $(aA)^{-1}$ must be $a^{-1}A^{-1}$, for *this latter satisfies a defining characteristic* of an inverse matrix.

(3) From the associative property of matrix multiplication, Theorem 2.6, part (1), we have

$$\begin{aligned}(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I \\ = B^{-1}B = B^{-1}IB = B^{-1}(A^{-1}A)B = (B^{-1}A^{-1})(AB)\end{aligned}$$

We are forced to conclude that $B^{-1}A^{-1}$ is the very inverse $(AB)^{-1}$.

(4) Exercise! ■

4.2 The Inverse of a 2×2 Matrix

Let us compute right here the formula for the inverse of an invertible 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(n, \mathbb{R})$$

Letting

$$A^{-1} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}(n, \mathbb{R})$$

what we need to do is find α, β, γ and δ in terms of a, b, c and d, using the only thing we have to go on, which is the defining identity

$$AA^{-1} = A^{-1}A = I$$

If we 'foil out' this identity, or expand it out, we'll probably get the formulas we want, but it might take a couple of moves to do it: First move is foil it out:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I = AA^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} a\alpha + b\gamma & a\beta + b\delta \\ c\alpha + d\gamma & c\beta + d\delta \end{pmatrix}$$

and similarly

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I = A^{-1}A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a\alpha + b\gamma & a\beta + b\delta \\ c\alpha + d\gamma & c\beta + d\gamma \end{pmatrix}$$

Each matrix equation provides us with a system of 4 equations in 4 unknowns (α, β, γ and δ),

$$\begin{aligned} a\alpha \quad + b\gamma \quad &= 1 \\ a\beta \quad + b\delta &= 0 \\ c\alpha \quad + d\gamma \quad &= 0 \\ c\beta \quad + d\delta &= 1 \end{aligned}$$

and

$$\begin{aligned} a\alpha + c\beta \quad &= 1 \\ b\alpha + d\beta \quad &= 0 \\ a\gamma + c\delta &= 0 \\ b\gamma + d\delta &= 1 \end{aligned}$$

i.e.

$$\left(\begin{array}{cccc|c} a & 0 & b & 0 & 1 \\ 0 & a & 0 & b & 0 \\ c & 0 & d & 0 & 0 \\ 0 & c & 0 & d & 1 \end{array}\right)$$

and

$$\left(\begin{array}{cccc|c} a & c & 0 & 0 & 1 \\ b & d & 0 & 0 & 0 \\ 0 & 0 & a & c & 0 \\ 0 & 0 & b & d & 1 \end{array}\right)$$

Wolfram Alpha (as well as my own scratchwork—*Exercise: perform these row-reductions for yourself, for practice!*) tells us that

$$\text{rref}\left(\begin{array}{cccc|c} a & 0 & b & 0 & 1 \\ 0 & a & 0 & b & 0 \\ c & 0 & d & 0 & 0 \\ 0 & c & 0 & d & 1 \end{array}\right) = \text{rref}\left(\begin{array}{cccc|c} a & c & 0 & 0 & 1 \\ b & d & 0 & 0 & 0 \\ 0 & 0 & a & c & 0 \\ 0 & 0 & b & d & 1 \end{array}\right) = \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & d/(ad-bc) \\ 0 & 1 & 0 & 0 & -b/(ad-bc) \\ 0 & 0 & 1 & 0 & -c/(ad-bc) \\ 0 & 0 & 0 & 1 & a/(ad-bc) \end{array}\right)$$

i.e.

$$\alpha = \frac{d}{ad-bc}; \qquad \beta = \frac{-b}{ad-bc};$$
$$\gamma = \frac{-c}{ad-bc}, \qquad \delta = \frac{a}{ad-bc},$$

from which we conclude that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Remark 4.7 All this is clearly only possible if the quantity $ad-bc$ is nonzero. This quantity is called the **determinant** of $A \in M_2(\mathbb{R})$, and it turns out to *completely characterize* the invertibility of $A \in M_2(\mathbb{R})$, as we will see in the section on determinants. *Any* 2×2 *matrix* A *is invertible iff* $\det A = ad - bc \neq 0$.

Remark 4.8 Here's a question to consider: Is there a *geometric* interpretation to $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$? Consider the *cross product* of the columns of A, viewed as lying in $\mathbb{R}^3$ with the z-coordinate 0:

$$\mathbf{a}_1 \times \mathbf{a}_2 = \begin{pmatrix} a \\ c \\ 0 \end{pmatrix} \times \begin{pmatrix} b \\ d \\ 0 \end{pmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a & c & 0 \\ b & d & 0 \end{vmatrix} = (ad-bc)\mathbf{k} = \begin{pmatrix} 0 \\ 0 \\ ad-bc \end{pmatrix}$$

and the length of this vector is

$$\|\mathbf{a}_1 \times \mathbf{a}_2\| = \|(ad-bc)\mathbf{k}\| = \|\langle 0, 0, ad-bc\rangle\| = \sqrt{0^2 + 0^2 + (ad-bc)^2} = |ad-bc|$$

This, we recall, is the area of the parallelogram spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$. ■

Exercise 4.9 We have actually only provided a *derivation* of the inverse of A, but to be sure, to *prove* that the result of our derivation is in fact the inverse, we need to crunch out the matrix product of A and A^{-1}:

$$A^{-1}A = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} ad-bc & bd-bd \\ -ac+ac & ad-bc \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$$

Compute AA^{-1} and verify that there, too, you get $AA^{-1} = I_2$. ■

5 Definition of Linear Transformations

Definition 5.1 We now arrive at the *abstract* **linear function** or **linear transformation** from $\mathbb{R}^n$ to $\mathbb{R}^m$. This is an opaque function

$$T : \mathbb{R}^n \to \mathbb{R}^m$$

and the only thing we know about it, by design, is that it is *characterized* by the following two properties:

(1) $T(c\mathbf{x}) = cT(\mathbf{x})$

(2) $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$

The set of all functions satisfying properties (1) and (2) is denoted

$$\boxed{\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \quad \text{or} \quad \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)}$$

depending on who you read, and is called the **space of linear transformations** (from $\mathbb{R}^n$ to $\mathbb{R}^m$). Our book doesn't use any notation for this, but we will use $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.[3] The reason for the word 'space' is that, like $M_{m,n}(\mathbb{R})$, $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ will turn out to be a *vector space*! ■

Notation 5.2 *When $n = m$, we have a simpler way to write $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ (and likewise for $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^n)$), namely*

$$\boxed{\mathcal{L}(\mathbb{R}^n) \quad \textit{or} \quad \mathrm{End}(\mathbb{R}^n)}$$

*This is similar to writing $M_n(\mathbb{R})$ instead of $M_{n,n}(\mathbb{R})$. Again, $\mathcal{L}$ reminds one of the linearity of a function $T \in \mathcal{L}(\mathbb{R}^n)$. Some people use the special term **linear operator** to describe such $T \in \mathcal{L}(\mathbb{R}^n)$. Whereas $\mathrm{End}(\mathbb{R}^n)$ stands for the catch-all algebraic term **endomorphism**, a term denoting an algebraic function having the same domain and codomain. We'll stick to the first notation.*

*Finally, we denote the group of all invertible linear transformations, also called the **general linear group**, by*

$$\boxed{\mathrm{GL}(\mathbb{R}^n)}$$

*but we remark that $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{GL}(\mathbb{R}^n)$, though similar and clearly related, are **different** groups. They'll turn out to be 'isomorphic', so 'basically the same' group, but the difference is one that needs to be understood.* ■

We pause here to catch our breath before attacking the next two topics,

(1) Bases, coordinates and representations.

(2) Determinant and trace.

---

[3] I am strongly tempted to use the other notation, $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$, because it is a little more standard, but that's only true at the higher levels. For our needs, $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ will serve its purpose by reminding us of the *linearity* of the functions with its $\mathcal{L}$. The other notation stands for 'homomorphisms,' a catch-all algebraic term that works in any algebraic category (not just vector spaces, but modules, groups, rings, etc.).