Bases, Coordinates and Representations

(3/27/19)

Alex Nita

Abstract

This section gets to the core of linear algebra: the separation of vectors and linear transformations from their *particular numerical realizations*, which, until now, we've largely taken for granted. We've taken them for granted because calculus and trigonometry were constructed *within* the standard Cartesian coordinate system for $\mathbb{R}^2$ and $\mathbb{R}^3$, and the issue was never raised. There *is* a way to construct a coordinate-free calculus, of course, just like there is a way to construct a coordinate-free linear algebra—the *abstract* way, *purely in terms of properties*—but I believe that misses the crucial point. We *want* coordinates. We need them to *apply* linear algebra to anything, because coordinates correspond to units of measurement forming a 'grid.'

Putting coordinates on an abstract vector space is like 'turning the lights on' in the dark. Except the abstract vector space is not *entirely* dark. The abstract vector space *has* structure, it's just that *the structure is defined purely in terms of properties*, not numbers. In the quintessentially modern Bourbaki style of math, where properties alone are considered, the *'abstract vector space over a field'* loses even that residual scalar-multiplicative role of numbers, replacing it with the *generic* scalar—the abstract number—the algebraic *field*. A field captures, by a clever use of (algebraic) properties, the commonalities of $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, finite fields $\mathbb{F}_p$, and other types of numbers.

The virtue of this approach is that many proofs become simplified—we no longer have to *compute* ad nauseam, we can just use properties, whose symbolic tidiness compresses or eliminates the mess of numbers. Another benefit is the *clarification* afforded by the compactness of the abstraction. Once we internalize and master the concepts, linear algebra becomes *transparent*, its structure unfolding effortlessly to the point where one can use its power with ease. This approach, incidentally, is that for the most part taken in MATH 2135, linear algebra for math majors.

On the other hand, what about those coordinates, so we can *use* linear algebra for something? They never go away, of course, and in fact their accomodation produces the most interesting part of this story: the rules for *representing* vectors and linear transformations, along with the the rules for *switching between different coordinate representations*. This is the entry point into eigenvalues and eigenvectors, as well as other questions of *optimal coordinates* for various purposes. Are there *canonical* or other special *forms* (read '*representation*') for linear transformations?

It turns out that certain types of matrices (or linear transformations) have corresponding special types of 'canonical' representations. This is the subject of the *spectral theorems*, which are related to *diagonalization*, *eigenvalues* and *eigenvectors*. For example, symmetric matrices (or symmetric linear operators) are always diagonalizable and have real eigenvectors. Their 'canonical form' is a diagonal matrix with real numbers down the diagonal. Skew-symmetric matrices are *almost*-diagonalizable—up to 2×2 rotation matrix blocks, if you don't ask for complex eigenvalues. This is *one* 'canonical' form of theirs.

But this is just the beginning. For we can, in one further move, come to view symmetric matrices as represention of *generalized dot products*, or *inner products*! Orthogonal matrices, whose first incarnations we saw in the 2×2 rotation and reflection matrices, turn out to *preserve the dot product*:

$$(A\mathbf{x}) \cdot (A\mathbf{x}) = \mathbf{x} \cdot \mathbf{x}$$

Not only that, but they're the *only* ones to do so—this dot-product-preservation is a property that *characterizes* orthogonal matrices. Since the length of a vector is the square root of its dot product with itself:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

orthogonal matrices preserve lengths:

$$\|A\mathbf{x}\| = \sqrt{(A\mathbf{x}) \cdot (A\mathbf{x})} = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \|\mathbf{x}\|$$

This is why orthogonal matrices are sometimes called *isometries*: because they keep lengths ('metries') the same ('iso'). Lengths (of the above type) and dot products are interchangeable/logically equivalent notions in $\mathbb{R}^n$ (hopefully we'll get to the proof of this).

One of the most successful methods of mathematics is 'taking things apart and putting them back together again,' with the 'putting things back together' part done rigorously, in a conceptually tight and unified way. The tactic is to separate out the tangle of concepts surrounding $\mathbb{R}^n$ (vector, linear transformation, basis, inner product, metric, eigenvalue), analyze each one separately, and then put them back together *intelligently, rigorously, in a conceptually unified way*. That will be our strategy! We begin with bases and coordinates.

1 Span and Linear Independence

Definition 1.1 A **linear combination** of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{R}^n$ is a sum

$$\boxed{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k} \tag{1.1}$$

where the a_i are *real numbers*, called the **coefficients**.[1] We say that the resulting vector $\mathbf{v} = \sum_{j=1}^{k} a_j\mathbf{v}_j$ ***is a linear combination of the*** $\mathbf{v}_i$. The number k here has nothing to do with n. It could be smaller, equal to, or much larger than n. It just has to be an integer greater than 0. ■

[1] Others, like the authors of our textbook, call them **weights**, a term one sees more often in statistics.

Example 1.2 Let $\mathbf{v}_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$, $\mathbf{v}_3 = \begin{pmatrix} 12 \\ 4 \end{pmatrix}$, $\mathbf{v}_4 = \begin{pmatrix} 70 \\ -22 \end{pmatrix}$ in $\mathbb{R}^2$. The following is a linear combination of the $\mathbf{v}_i$:

$$\begin{aligned}(-3)\mathbf{v}_1 + 2\mathbf{v}_2 + \frac{1}{4}\mathbf{v}_3 + 0\mathbf{v}_4 &= -3\begin{pmatrix} -1 \\ 2 \end{pmatrix} + 2\begin{pmatrix} 3 \\ 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 12 \\ 4 \end{pmatrix} + 0\begin{pmatrix} 70 \\ -22 \end{pmatrix} \\ &= \begin{pmatrix} 3+6+3+0 \\ -6+0+1+0 \end{pmatrix} \\ &= \begin{pmatrix} 12 \\ -5 \end{pmatrix}\end{aligned}$$

We say that *the vector* $\begin{pmatrix} 12 \\ -5 \end{pmatrix}$ *is a linear combination of the vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_4 \in \mathbb{R}^2$. ■

Definition 1.3 We can generalize this to any (*potentially infinite*) collection of vectors V in $\mathbb{R}^n$, for example, the unit disk $V = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \leq 1\}$. It goes like this: a **linear combination** of vectors in V is by definition a *finite* sum of vectors in V, exactly as in (1.1), no matter that V may contain infinitely many vectors in itself. *We only ever add finitely many vectors.* ■

Definition 1.4 Define the **span** of a collection of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ to be *set of all linear combinations of the* $\mathbf{v}_i$, and denote it

$$\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \stackrel{\text{def}}{=} \left\{ \sum_{j=1}^{k} a_j \mathbf{v}_j \;\middle|\; a_j \in \mathbb{R},\ k \in \mathbb{N},\ j = 1, \ldots, k \right\}$$

This generalizes to *any*, even infinite-sized sets $V \subseteq \mathbb{R}^n$ of vectors in $\mathbb{R}^n$, by adding one statement about V in the property list on the right:

$$\operatorname{span}(V) \stackrel{\text{def}}{=} \left\{ \sum_{j=1}^{k} a_j \mathbf{v}_j \;\middle|\; a_i \in \mathbb{R},\ \mathbf{v}_j \in V,\ k \in \mathbb{N},\ j = 1, \ldots, k \right\}$$

■

To round out an idea in math, what we do is literally just *write down* the set of all things of the kind we have in mind—we write it as their defining property (e.g. $\{x \mid x$ has property $P\}$). Then we give that set a name. That way we don't have to go around calling it 'the set of all things of type such-and-such,' especially when 'such-and-such' is itself a mouthful. So here, with the span.

Definition 1.5 A subset U of $\mathbb{R}^n$ is called a **subspace** if it contains $\mathbf{0}$ and is *closed* under addition and scalar multiplication

(1) $\mathbf{0} \in U$

(2) $\mathbf{u}, \mathbf{v} \in U \implies \mathbf{u} + \mathbf{v} \in U$

(3) $\mathbf{u} \in U$ and $c \in \mathbb{R} \implies c\mathbf{u} \in U$

We say U is a **proper subspace** if it is a proper (i.e. strictly smaller) subset of $\mathbb{R}^n$. ■

Proposition 1.6 *Let V be any collection of vectors in $\mathbb{R}^n$ (a sub__set__ of $\mathbb{R}^n$). Then,* span(V) *is a sub__space__ of $\mathbb{R}^n$.*

Proof: Since $\text{span}(V) = \left\{\sum_{j=1}^{k} a_j\mathbf{v}_j \,\middle|\, a_j \in \mathbb{R},\ \mathbf{v}_j \in V,\ k \in \mathbb{N}\right\}$, we clearly have $\mathbf{0} = \sum_{j=1}^{k} 0\mathbf{v}_j$ for any $\mathbf{v}_j \in V$. Now suppose

$$\mathbf{u} = \sum_{j=1}^{\ell} a_j\mathbf{u}_j, \quad \mathbf{v} = \sum_{j=1}^{k} b_j\mathbf{v}_j \in \text{span}(V)$$

where the $\mathbf{u}_i$ and $\mathbf{v}_j$ are in V. We do not assume any of these are the same, but if some of them are, let us relabel these to keep track: call the $\mathbf{u}$-duplicates $\mathbf{w}_{\ell-r+s}$ and the $\mathbf{v}$-duplicates $\mathbf{w}_{k-r+s}$, where $s = 1, \ldots r$, and let's just tidy up by re-indexing everything so that the $\mathbf{w}$'s appear last in the sums:

All this work was done to write the *resulting* linear combination (1.2) as a *single* sum, so as to have it in exactly the form that Definition 1.4 requires. It was a bit of *fussing* over details, because when you add two finite sums, such as those for $\mathbf{u}$ and $\mathbf{v}$, you still get a finite sum. You just have to regroup and relabel things—cosmetic touch-ups—to get it to look like a single sum.

$$\mathbf{u} = \sum_{i=1}^{\ell-r} a_i\mathbf{u}_i + \sum_{s=1}^{r} a_{\ell-r+s}\mathbf{w}_{k-r+s},$$
$$\mathbf{v} = \sum_{j=1}^{k-r} b_j\mathbf{v}_j + \sum_{s=1}^{r} b_{k-r+s}\mathbf{w}_{k-r+s}$$

The sums $\sum_{i=1}^{\ell-r} b_i\mathbf{u}_i$ and $\sum_{j=1}^{k-r} a_j\mathbf{v}_j$ are each linear combinations of different vectors in V, so they *and their sum* lie in span(V), whereas the $\mathbf{w}$-parts require associativity to tidy-up the sum first:

$$\sum_{s=1}^{r} a_{\ell-r+s}\mathbf{w}_{k-r+s} + \sum_{s=1}^{r} b_{k-r+s}\mathbf{w}_{k-r+s} = \underbrace{\sum_{s=1}^{r}(a_{\ell-r+s} + b_{k-r+s})\mathbf{w}_s}_{\text{re-indexing the } \mathbf{w}\text{'s again}}$$

This, too, now lies in span(V), so the full sum

$$\mathbf{u}+\mathbf{v} = \sum_{i=1}^{\ell-r} b_i\mathbf{u}_i + \sum_{j=1}^{k-r} a_j\mathbf{v}_j + \sum_{s=1}^{r}(a_{\ell-r+s} + b_{k-r+s})\mathbf{w}_s = \underbrace{\sum_{i=1}^{\ell+k-r} a_i\mathbf{v}_i}_{\substack{\text{relabeling the}\\ \text{vectors and scalars}}} \tag{1.2}$$

is now seen to be a linear combination of vectors in V, hence to lie in span(V).

Similarly, if $c \in \mathbb{R}$, distributivity and associativity of scalar multiplication give

$$c\mathbf{u} = c\sum_{i=1}^{\ell} a_i\mathbf{u}_i = \sum_{i=1}^{\ell}(ca_i)\mathbf{u}_i \in \text{span}(V) \qquad \blacksquare$$

Example 1.7 We now have several examples of subspaces of $\mathbb{R}^n$:

(1) $\text{span}(V)$ for any $V \subseteq \mathbb{R}^n$

(2) $\text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ is a special case of (1), where $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$

(3) $N(A)$ for any $A \in M_{m,n}(\mathbb{R})$ is a subspace of $\mathbb{R}^n$

(4) $R(A)$ for any $A \in M_{m,n}(\mathbb{R})$ is a subspace of $\mathbb{R}^m$

We could include others, such as

(5) The **trivial subspace** $\{\mathbf{0}\}$

(6) The **total space** $\mathbb{R}^n$ ■

What other examples of subspaces do we have?

Exercise 1.8 Show that the trivial subspace $\{\mathbf{0}\}$ is in fact a subspace. ■

Exercise 1.9 If U and V are subspaces of $\mathbb{R}^n$, show that the following are also subspaces:

(1) $U \cap V = \{\mathbf{u} \in \mathbb{R}^n \mid \mathbf{u} \in U \text{ and } \mathbf{u} \in V\}$

(2) $U + V = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in U \text{ and } \mathbf{v} \in V\}$ ■

Repeat the above exercise, but now with finitely many subspaces U_i.

Exercise 1.10 If $U_1, \ldots U_\ell$ are subspaces of $\mathbb{R}^n$, show that the following are also subspaces:

(1) $\displaystyle\bigcap_{i=1}^{\ell} U_i \;=\; U_1 \cap U_2 \cap \cdots \cap U_\ell = \{\mathbf{u} \in \mathbb{R}^n \mid \mathbf{u} \in U_i \text{ for all } i = 1, \ldots, \ell\}$

(2) $\displaystyle\sum_{i=1}^{\ell} U_i \;=\; U_1 + U_2 + \cdots + U_\ell = \{\mathbf{u}_1 + \cdots + \mathbf{u}_\ell \mid \mathbf{u}_i \in U_i \text{ for each } i = 1, \ldots, \ell\}$ ■

Definition 1.11 If U and V are subspaces of $\mathbb{R}^n$, then the previous exercises show

$$\boxed{U + V}$$

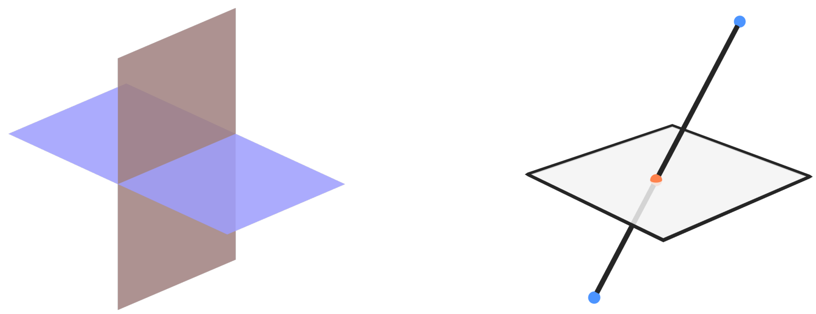
is also a subspace. This subspace is called the **sum** of the subspaces. If additionally U and V intersect trivially, $U \cap V = \{\mathbf{0}\}$, the sum is called a **direct sum**, and denoted

$$\boxed{U \oplus V}$$

This idea can be extended to any finite sum: if $U_1, \ldots,\ U_k$ are subspaces of $\mathbb{R}^n$ pairwise trivially intersecting, $U_i \cap U_j = \{\mathbf{0}\}$, their direct sum is denoted

$$\boxed{\bigoplus_{i=1}^{k} U_i \;=\; U_1 \oplus U_2 \oplus \cdots \oplus U_k}$$ ■

Example 1.12 Two planes $\mathcal{P}$ and $\mathcal{Q}$ in $\mathbb{R}^3$ intersect in a line ℓ, a nontrivial subspace, while a plane and a line *off* the plane intesect trivially (at the origin):



(a) $\mathcal{P}+\mathcal{Q}=\mathbb{R}^3$, but $\mathcal{P}\cap\mathcal{Q}=\ell\neq\{\mathbf{0}\}$

(b) $\mathcal{P}+\ell=\mathbb{R}^3$, *and* $\mathcal{P}\cap\ell=\{\mathbf{0}\}$

Figure 1.1: A sum which is *not* direct, $\mathcal{P}+\mathcal{Q}=\mathbb{R}^3$, vs a sum which *is* direct, $\mathcal{P}\oplus\ell=\mathbb{R}^3$.

Example 1.13 $\mathbb{R}^n$ may be decomposed in various ways as a direct sum, for example

$$\begin{aligned}\mathbb{R}^n &= \bigoplus_{i=1}^{n}\operatorname{span}(\mathbf{e}_i)\\ &= \operatorname{span}(\mathbf{e}_1)\oplus\operatorname{span}(\mathbf{e}_2)\oplus\cdots\oplus\operatorname{span}(\mathbf{e}_n)\end{aligned}$$

or

$$\mathbb{R}^n = \operatorname{span}(\mathbf{e}_1,\ldots,\mathbf{e}_k)\oplus\operatorname{span}(\mathbf{e}_{k+1},\ldots,\mathbf{e}_n)$$

Or how about this one:

$$\mathbb{R}^3 = \operatorname{span}\begin{pmatrix}1\\1\\0\end{pmatrix}\oplus\operatorname{span}\begin{pmatrix}1\\1\\1\end{pmatrix}\oplus\operatorname{span}\begin{pmatrix}1\\-1\\0\end{pmatrix}$$

■

Remark 1.14 If instead of using the standard basis vectors $\mathbf{e}_i$, we use instead, say, $\mathbf{b}_j=\sum_{i=1}^{j}\mathbf{e}_i$ (i.e. $\mathbf{b}_1=\mathbf{e}_1$, then $\mathbf{b}_2=\mathbf{e}_1+\mathbf{e}_2$, etc.), then we'll have

$$\mathbb{R}^n=\sum_{i=1}^{n}\operatorname{span}(\mathbf{b}_i)$$

but the sum will not be direct. For example, $\operatorname{span}(\mathbf{b}_1)\cap\operatorname{span}(\mathbf{b}_2)=\operatorname{span}(\mathbf{e}_1)$, which is not $\{\mathbf{0}\}$. ■

Definition 1.15 Let $A \in M_{m,n}(\mathbb{R})$, and let $\vec{A}_i$ denote the rows of A, $i = 1, \dots, m$. Define the **row space** of A to be the span of the rows,

$$\boxed{\mathrm{Row}(A) := \mathrm{span}(\vec{A}_1, \dots, \vec{A}_m)}$$

This is a subspace of $\mathbb{R}^n$, by Proposition 1.6, and its importance can be seen in the next example—a theorem in fact. We can decompose $\mathbb{R}^n$, the domain of A, into a direct sum of null space and row space. ■

Example 1.16 As another example, if $A \in M_{m,n}(\mathbb{R})$, then both $N(A)$ and $\mathrm{Row}(A)$ are subspaces of $\mathbb{R}^n$. Moreover, as you will show in Exercise 1.18, they intersect trivially, $N(A) \cap \mathrm{Row} = \{\mathbf{0}\}$, so their sum is direct, and is another subspace of $\mathbb{R}^n$. In fact, we will prove below that

$$\boxed{\mathbb{R}^n = N(A) \oplus \mathrm{Row}(A)}$$

■

Example 1.17 To illustrate the previous example with some numbers, conisder

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \dashrightarrow B = \mathrm{rref}(A) = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

I claim that $\mathrm{Row}(A) = \mathrm{Row}(B)$, i.e.

$$\mathrm{span}(\langle 1,2,3\rangle, \langle 4,5,6\rangle, \langle 7,8,9\rangle) = \mathrm{span}(\langle 1,0,-2\rangle, \langle 0,1,2\rangle)$$

but you must wait just a little for the proof—it will come shortly. Let's just use the result. We only need one more thing, $N(A)$. Parametrize the free variable z, etc.:

$$z = t, \implies x = z = t,\ y = -2t \implies \mathbf{x} = \begin{pmatrix} t \\ -2t \\ t \end{pmatrix} = t \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

or in other words

$$N(A) = \mathrm{span}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

Consequently,

$$\boxed{\begin{aligned} \mathbb{R}^3 &= N(A) \oplus \mathrm{Row}(A) \\ &= N(A) \oplus \mathrm{Row}(B) \\ &= \mathrm{span}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \oplus \mathrm{span}\left(\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}\right) \end{aligned}}$$

■

Exercise 1.18 Let $A \in M_{m,n}(\mathbb{R})$. Use the dot-product definition of matrix-vector multiplication to show that the subspaces $N(A)$ and $\text{Row}(A) := \text{span}(\vec{A}_1, \ldots, \vec{A}_m)$ are *orthogonal*, in the sense that any $\mathbf{u} \in N(A)$ and $\mathbf{v} \in \text{Row}(A)$ are orthogonal, $\mathbf{u} \cdot \mathbf{v} = 0$. From this, deduce that

$$N(A) \cap \text{Row}(A) = \{\mathbf{0}\}$$

(Hint: if $\mathbf{u} \cdot \mathbf{v} = 0$, then see what happens to any vector $\mathbf{x} = \text{span}(\mathbf{u}) \cap \text{span}(\mathbf{v})$. What is $\|\mathbf{x}\|$?) From this, we conclude that the sum is direct,

$$N(A) \oplus \text{Row}(A)$$

<u>Note</u>: we don't know that this direct sum is equal *to* $\mathbb{R}^n$, *but only that it's a* subspace *of it.* For example,

$$\text{span}(\mathbf{e}_1) \oplus \text{span}(\mathbf{e}_2)$$

is a *proper* subspace of $\mathbb{R}^3$, so does not *equal* $\mathbb{R}^3$. To show strict *equality* takes an addtional argument—a *dimension* argument: the dimensions of $N(A)$ and $\text{Row}(A)$ need to add up to the dimension, n, of $\mathbb{R}^n$. This will be shown below, in the Rank-Nullity Theorem. ■

Exercise 1.19 Let U be the solution set of a system $A\mathbf{x} = \mathbf{0}$ and V be the solution set of $B\mathbf{x} = \mathbf{0}$, where $A, B \in M_{m,n}(\mathbb{R})$. Show that the set of *common solutions* $U \cap V$ is a subspace of $\mathbb{R}^n$. ■

Exercise 1.20 Give an example to show that, in general, $U \cup V$ is not a subspace, if U and V are proper subspaces. ■

Exercise 1.21 This is a very difficult exercise, but it's a solid test of math skills: Show that $\mathbb{R}^n$ is not a finite union of proper subspaces. Maybe consult Math Stack Exchange for some hints, but not until you bang your head against it. ■

Definition 1.22 A collection of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is called **linearly independent** if, whenever we *represent* the zero vector as a linear combination of them

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$$

we are *forced* to conclude that the coefficients are *all zero*:

$$a_1 = a_2 = \cdots = a_k = 0$$

Expressed succinctly, this says:

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = \mathbf{0} \quad \Longrightarrow \quad a_1 = a_2 = \cdots = a_k = 0$$

where the arrow means 'implies.' Another way of saying this is, *'The only representation of the zero vector is the **trivial representation**,'*

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_k = \mathbf{0}$$

which, of course, we always have (by Corollary 1.12, '*Points and Vectors in* $\mathbb{R}^n$').

If there exists *even one* non-trivial representation of $\mathbf{0}$, we say that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are **linearly dependent**. ■

The notion of *representing* a vector will be explained below, when we get to bases. In this case, the representation is of the zero vector $\mathbf{0}$, and it is *in the basis* $\{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ for the subspace $U = \text{span}(\mathbf{v}_1, \ldots \mathbf{v}_k)$.

Exercise 1.23 Show that the three vectors we got in Example 1.17,

$$\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

are linearly independent in $\mathbb{R}^3$. ■

Exercise 1.24 Show that the rows of

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

are linearly dependent, by finding a nontrivial representation of $\mathbf{0} = \langle 0, 0, 0 \rangle$ in their span. ■

Exercise 1.25 Show that the four vectors in Example 1.2 are linearly dependent in $\mathbb{R}^2$, by finding a nontrivial representation of $\mathbf{0} = \binom{0}{0}$ in their span. ■

Theorem 1.26 (Characterization of Linear Independence 1) *Vectors* $\mathbf{v}_1, \ldots,$ $\mathbf{v}_k$ *in* $\mathbb{R}^n$ *are linearly independent iff none of them is a linear combination of the remainder, i.e.*

$$\begin{array}{c} \mathbf{v}_1, \ldots, \mathbf{v}_k \ \textit{ are linearly independent} \\ \Longleftrightarrow \\ \textit{no } \mathbf{v}_i \textit{ lies in } \text{span}\big(\mathbf{v}_1 \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k\big) \end{array}$$

Proof: Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ are linearly independent. This means that whenever $\sum_{j=1}^k a_i \mathbf{v}_i = \mathbf{0}$, we must have all $a_i = 0$. If, therefore, we erroneously assume that it may be possible for one of the vectors to be in the span of the others,

$$\mathbf{v}_i \in \text{span}\big(\mathbf{v}_1 \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k\big)$$

that is, that

$$\mathbf{v}_i = a_1\mathbf{v}_1 + \cdots + a_{i-1}\mathbf{v}_{i-1} + a_{i+1}\mathbf{v}_{i+1} + \cdots + a_k\mathbf{v}_k$$

then we're one algebraic move away from a contradiction: subtract $\mathbf{v}_i$ from both sides, and you have a nontrivial representation of $\mathbf{0}$:

$$\mathbf{0} = \big(a_1\mathbf{v}_1 + \cdots + a_{i-1}\mathbf{v}_{i-1} + a_{i+1}\mathbf{v}_{i+1} + \cdots + a_k\mathbf{v}_k\big) + (-1)\mathbf{v}_i$$

because of that -1. This shows that if the vectors are linearly independent, none can be in the span of the others.

Now for the reverse direction: If none of the $\mathbf{v}_i$ is in the span of the others, then suppose

$$\sum_{i=1}^{k} a_i\mathbf{v}_i = \mathbf{0}$$

If you imagine, erroneously, that it might be possible for the linear combination to be nontrivial, then you are setting yourself up. It's like negligently giving away your queen in chess, for by supposing there is *even one* nonzero a_i, then you open yourself up to algebraic defeat. Divide through by a_i, then solve for $\mathbf{v}_i$ and undeceive yourself, friend!

$$\mathbf{v}_i = \sum_{i\neq j}\Big(-\frac{a_j}{a_i}\Big)\mathbf{v}_j$$

You'll have to admit that the above equation displays $\mathbf{v}_i$ as a member of the span of the remainder—there's the error. We conclude that if none of the vectors is in the span of the others, then they must be linearly independent. ■

Exercise 1.27 Let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ be any vectors, not necessarily linearly independent. Show that

$$\operatorname{span}\big(\mathbf{v}_1, \ldots, \mathbf{v}_k\big) = \operatorname{span}(\mathbf{v}_1) + \cdots + \operatorname{span}(\mathbf{v}_k)$$

Exercise 1.28 Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Show that the sum in $\operatorname{span}(\mathbf{u}, \mathbf{v}) = \operatorname{span}(\mathbf{u}) + \operatorname{span}(\mathbf{v})$ is direct:

$$\operatorname{span}(\mathbf{u}) \oplus \operatorname{span}(\mathbf{v}) = \operatorname{span}(\mathbf{u}, \mathbf{v})$$

iff $\mathbf{u}$ and $\mathbf{v}$ are linearly independent. Is this subspace equal to $\operatorname{span}(\mathbf{u} + \mathbf{v})$? ■

Remark 1.29 The previous exercise extends to any finite number of vectors, not just two. The proof is essentially that of the previous exercise, though: to show that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{R}^n$ are linearly independent iff

$$\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \bigoplus_{i=1}^{k} \operatorname{span}(\mathbf{v}_i)$$

all you need to actually show is that the sum is direct, but to show this you just do it for any pair of spans, that is, you just need to show $\operatorname{span}(\mathbf{v}_i) \cap \operatorname{span}(\mathbf{v}_j) = \{\mathbf{0}\}$, which is just the previous exercise with decorations (i and j). ■

2 One-to-One Matrix Functions Preserve Linear Independence

Let us reap an immediate reward for our efforts. Having introduced linear independence as a new concept, it is natural to ask how this concept interacts with the others we introduced before. Well, the most immediate is this one: one-to-one linear functions (which includes matrix functions by Theorem 2.4, 'Matrices and Linear Transformations') are the ones which *preserve* linear independence. We prove the matrix version here, and await the general one while we sort out the coordinate representations of linear transformations.

Theorem 2.1 (One-to-One Matrices Preserve Linear Independence)
Let $A \in M_{m,n}(\mathbb{R})$ be any matrix, thought of as a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then, A is one-to-one iff it carries linearly independent sets $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ to linearly independent sets $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\}$ in $\mathbb{R}^m$.

Proof: Let $A \in M_{m,n}(\mathbb{R})$. Suppose first that A is one-to-one, and take a look at what happens when you feed a linearly independent set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ into it: is it possible that the coefficients a_i in the representation of $\mathbf{0}$,

$$a_1(A\mathbf{v}_1) + \cdots + a_k(A\mathbf{v}_k) = \mathbf{0}$$

are nonzero? No way: use the two properties of Theorem 2.4 in 'Matrices and Linear Transformations' to distribute A,

$$\begin{aligned} A(a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k) &= A(a_1\mathbf{v}_1) + \cdots + A(a_k\mathbf{v}_k) \\ &= a_1(A\mathbf{v}_1) + \cdots + a_k(A\mathbf{v}_k) \\ &= \mathbf{0} \qquad \textit{(by the assumption above)} \end{aligned}$$

and observe what this means: $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k \in N(A)$. But A is one-to-one, so by Porposition 3.9 $N(A) = \{\mathbf{0}\}$—therefore

$$a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$$

Aha! Now we're in business: the $\mathbf{v}_i$ are linearly independent in $\mathbb{R}^n$, by assumption, so the a_i *have to be* 0. This shows that $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\}$ is linearly independent in $\mathbb{R}^m$.

For the converse, now: If $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\}$ is linearly independent whenever $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is, then let's show $N(A) = \{\mathbf{0}\}$ and use Porposition 3.9 in 'Matrices and Linear Transformations' to get A's injectivity: Suppose $\mathbf{x} \in N(A)$, so that $A\mathbf{x} = \mathbf{0}$. Writing $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, and noting that $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is linearly independent (see Exercise 3.2 below), we get that

$$\begin{aligned} \mathbf{0} &= A\mathbf{x} \qquad (\mathbf{x} \in N(A) \textit{ by the assumption}) \\ &= x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \qquad \textit{(Definition 1.10)} \end{aligned}$$

But now we cash in our assumption: If $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is inearly independent, then so is $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \{A\mathbf{e}_1, \ldots, A\mathbf{e}_n\}$. This means the x_i above all have to be 0, which means $\mathbf{x} = 0$, showing that $N(A) = \{\mathbf{0}\}$. By Porposition 3.9 ('Matrices and Linear Transformations') this means A is one-to-one. ■

3 Basis and Dimension

Definition 3.1 An *unordered* collection of vectors

$$\beta = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$$

in $\mathbb{R}^n$ which is ***both***

(1) ***linearly independent***, and

(2) ***spans*** $\mathbb{R}^n$

is called a **basis** for $\mathbb{R}^n$. The '*unordered*' part means the indexing can be changed for convenience if desired. An **ordered basis** for $\mathbb{R}^n$, on the other hand, keeps the indexing fixed, and is denoted with round brackets:

$$\beta = \big(\mathbf{b}_1, \dots, \mathbf{b}_k\big)$$ ■

If we are concerned about ***orientation***, which is a generalized way of distinguishing 'left-handed' from 'right-handed,' then we need an *ordered* basis. To tell whether a given basis is 'right-handed' or 'left-handed' is determined by comparison with other bases: the *determinant* can be used to measure the 'orientation difference': If the change-of-coordinates matrix $M_{\beta,\gamma}$, which changes β-coordinates to γ-coordinates has positive determinant, $\det M_{\beta,\gamma} > 0$, then we say β and γ 'have the same orientation.' The natural choice to compare with is the *standard basis* σ (see Example 3.2 to the left), which we declare to have 'positive orientation.' If $\det M_{\beta,\sigma} > 0$, we'll say β is 'positively oriented.'

Example 3.2 The **standard basis** for $\mathbb{R}^n$ is the *ordered basis*

$$\sigma = \big(\mathbf{e}_1, \dots, \mathbf{e}_n\big) = \left(\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \dots\dots, \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix} \right)$$

It is certainly linearly independent, because the matrix equation $I_n\mathbf{x} = \mathbf{0}$ has only the trivial solution, which can be seen directly (see Definition 1.10, def 1, '*Matrices and Linear Transformations*'):

$$\mathbf{0} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n = \begin{pmatrix}1 & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & 1\end{pmatrix}\begin{pmatrix}a_1\\ \vdots\\ a_n\end{pmatrix} = I_n\mathbf{a} = \mathbf{a}$$

Thus, all $a_i = 0$, and we conclude that σ is linearly independent. It is also spanning, because any $\mathbf{x}$ lies in its span (see Definition 1.14, '*Points and Vectors in* $\mathbb{R}^n$'):

$$\mathbf{x} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n \in \operatorname{span}\big(\mathbf{e}_1, \dots, \mathbf{e}_n\big)$$ ■

Exercise 3.3 In Exercise 1.23 you showed that the three vectors

$$\begin{pmatrix}1\\-2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\-2\end{pmatrix}, \begin{pmatrix}0\\1\\2\end{pmatrix}$$

are linearly independent. Show that they also span $\mathbb{R}^3$, and so form a basis. ■

Observation 3.4 For any $\mathbf{v} \in \mathbb{R}^n$ we have $0\mathbf{v} = \mathbf{0}$.

Observation 3.5 For any $c \in \mathbb{R}$ we have $c\mathbf{0} = \mathbf{0}$.

Observation 3.6 Let $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$. If $c\mathbf{v} = \mathbf{0}$, then either $c = 0$ or $\mathbf{v} = \mathbf{0}$.

Theorem 3.7 (Equivalent Characterizations of Bases) *Let $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a subset of $\mathbb{R}^n$. The following are logically equivalent statements about β:*

(1) *β is a **basis** for $\mathbb{R}^n$.*

(2) *Every nonzero vector $\mathbf{v} \in \mathbb{R}^n$ has a **unique representation** in β, meaning*

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$$

*for unique $a_i \in \mathbb{R}$, which are called β-**coordinates**.*

(3) $\operatorname{span}(\beta) = \mathbb{R}^n$ *but no $\mathbf{v}_i$ lies in the span of the remainder.*

(4) *β is a **minimal spanning set**, meaning* $\operatorname{span}(\beta) = \mathbb{R}^n$ *but no proper subset $\tilde{\beta}$ of β spans $\mathbb{R}^n$.*

(5) *β is a **maximal linearly independent set**, meaning β is linearly independent, but no strictly larger extension γ of β, i.e. $\beta \subset \gamma$, is linearly independent.*

All of this is true more generally for any nontrivial subspace V of $\mathbb{R}^n$: any basis β for V is equivalently characterized by any of the other five conditions.

This says that $\operatorname{span}(\beta) = \mathbb{R}^n$, and *uniquely* so for *all* vectors: nonzero vectors have unique coordinate representations in β, and the zero vector has only the trivial representation in β, by its linear independence.

These last two points say that if you *remove* even one vector $\mathbf{v}_i$ from β you'll lose spanning, but if you *add even one* vector $\mathbf{v}$ to β you'll lose linear independence. The perfect balance, where you have both, is what *characterizes* a basis.

Exercise 3.8 Show that a collection $\beta = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ of vectors is a basis for $\mathbb{R}^n$ iff $\mathbb{R}^n$ decomposes into a direct sum

$$\mathbb{R}^n = \operatorname{span}(\mathbf{b}_1) \oplus \cdots \oplus \operatorname{span}(\mathbf{b}_k)$$

You may assume that $k = n$ here, though this is not proved until Theorem 3.13. ■

Remark 3.9 We will show below, Theorem 3.13, that any basis β for *any subspace* V of $\mathbb{R}^n$ will have a fixed size, k, which we call the dimension of V. Moreover, by the Replacement Theorem 3.10, we must conclude that $1 \leq \dim V \leq \dim \mathbb{R}^n = n$. ■

Proof: Step 1: We first prove the *equivalence of (1) and (2)*: Suppose characterization (1), that $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for $\mathbb{R}^n$. This means β is linearly independent and spans $\mathbb{R}^n$. Given an arbitrary vector $\mathbf{v} \in \mathbb{R}^n = \operatorname{span}(\beta)$, we know that $\mathbf{v}$ is in the span of β, so

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$$

Here's why bases are *exactly* the things you need to *uniquely* and *unambiguously* give the coordinates of any other vector. This is why coordinates are *meaningful only in terms of* bases. And, importantly, *bases are easy to find*, as we'll see.

for some $a_i \in \mathbb{R}$. What we need to show is that $\mathbf{v}$ is *uniquely* represented in these coordinates a_i. This must naturally involve the other aspect of β, its linear independence—anyway, how else are we going to get uniqueness for $\mathbf{v}$ if not from the uniqueness for $\mathbf{0}$, naturally via a little thing called subtraction? Well, let's do our job and suppose there were two representations of $\mathbf{v}$, the one above and perhaps another one,

$$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k$$

We now use the inexorable logic of algebra to crank our the desired conclusion, that all $a_i = b_i$, meaning there had only been *one* representation this whole time. Here we go: Given those two hypothetical representations, we merely subtract, group things, then hit the result with the linear independence assumption for a home run:

$$\begin{aligned} \mathbf{0} &= \mathbf{v} - \mathbf{v} \\ &= (a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k) - (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k) \\ &= (a_1 - b_1)\mathbf{v}_1 + \cdots + (a_k - b_k)\mathbf{v}_k \end{aligned}$$

This exhibits $\mathbf{0}$ represented in β, which, by its linearly independence, forces

$$a_1 - b_1 \;=\; a_2 - b_2 \;=\; \cdots \;=\; a_k - b_k \;=\; 0$$

and therefore

$$a_1 = b_1, \quad a_2 = b_2, \;\cdots,\; a_k = b_k$$

Thus, (1) implies (2).

Now let's prove the reverse, that (2) implies (1): Suppose characterization (2), that β gives unique representations to all nonzero vectors $\mathbf{v} \in \mathbb{R}^n$. Well, then β definitely spans $\mathbb{R}^n$—not only is every vector $\mathbf{v}$ in β's span, but it's *uniquely* so. Well, this uniqueness must also give linear independence, which is the just the uniqueness of the β-representation of $\mathbf{0}$. Suppose, then, that $\mathbf{0}$ has a possibly different β-representation,

$$\mathbf{0} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$$

Should we foolishly suppose that any of the coefficients a_i are nonzero, we'd get slapped with the contradiction on uniqueness of nonzero β-representations, like so: if, say, $a_i \neq 0$, then we can solve for $a_i\mathbf{v}_i$ and divide through by a_i,

$$\mathbf{v}_i = \sum_{j \neq i} \left(-\frac{a_j}{a_i}\right)\mathbf{v}_j$$

which makes it look like $\mathbf{v}_i$, a nonzero vector, has two different β-representations. That's a no-no, so we withdraw any such foolish assumption. We conclude that all $a_i = 0$. Then β is linearly independent, too, and so forms a basis. This shows that (2) implies (1), and finishes the equivalence of (1) and (2).

Step 2: We now prove the *equivalence of (1) and (3)*: Suppose characterization (1), that β is a basis. Well, it's linearly independent, so Theorem 1.26 applies to give that no $\mathbf{v}_i$ is in the span of the remainder. Since β is a basis, it also spans $\mathbb{R}^n$, which is

This equivalence is a pedantic repetition of Theorem 1.26, but with the addition of the *spanning* property, which concerns bases.

the other half of (3). Now suppose characterization (3) of β. Then β spans $\mathbb{R}^n$, and is linearly independent by Theorem 1.26 again. It is a basis, and (3) implies (1).

Step 3: We show the *equivalence of (1) and (4)*: Suppose characterization (1), that β is a basis. To see that no *proper* subset of β could possibly span $\mathbb{R}^n$, we have to notice the problem with supposing otherwise. If a proper subset $\tilde{\beta}$ of β somehow *did* span $\mathbb{R}^n$, then any missing vector $\mathbf{v}_i$ left out of β by $\tilde{\beta}$ would lie in the span of $\tilde{\beta}$, which spans all of $\mathbb{R}^n$, in flagrant violation of Theorem 3.7, according to which no linearly independent set such as β would allow any of its members to lie in the span of the remainder. We conclude that (1) implies (4).

This equivalence is important, because, along with the next equivalence, it will allow us to *count* the number of basis vectors and be *sure* we've got the right *number* necessary for the job. This number is the **dimension** of the vector space $\mathbb{R}^n$. Later, we'll show that any subspace, too, has a basis, and all bases have the *same number* of vectors, namely the *dimension* of the subspace. But this idea extends further, into the absolutely general idea of an abstract vector space. To be continued...

Now for the reverse, (4) implies (1). We suppose characterization (4), that β is minimal spanning. Then β must be linearly independent, for otherwise, were β *not* linearly independent, then Theorem 1.26 would force one of the $\mathbf{v}_i$ in β to lie in the span of the remainder, whence we could simply remove it and *still* have a spanning set[2]—a contradiction of minimality in spanning. We conclude that β is linearly independent. Since it's already spanning, we're done, β is a basis. This completes the equivalence of (1) and (4).

Step 4: In this last step, we show the *equivalence of (1) and (5)*: Suppose characterization (1), that β is a basis. We play the fool again and suppose β somehow *could* fail characterization (5), so that β *might not* be maximally linearly independent. Well, that means there's a strictly larger extension γ of β which is *still* linearly independent. Even *just one* extra vector $\mathbf{v}$ added to β,

$$\gamma = \beta \cup \{\mathbf{v}\} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}\}$$

would run afoul of Theorem 1.26, because $\mathbf{v}$ would be in the span of β, the remainder of the vectors γ after removing $\mathbf{v}$, which would kill the linear independence of γ. Oops! We quit playing the fool and admit (1) implies (5).

Now for the converse, (5) implies (1): Suppose β is characterized by being *maximally linearly independent*. Well, that means β is linearly independent, I guess. If we can show β spans $\mathbb{R}^n$, then we'll be sure it's a basis. We play the fool and try to see what happens if β *didn't* span $\mathbb{R}^n$. Then there'd be a vector $\mathbf{v} \in \mathbb{R}^n$ that isn't in $\text{span}(\beta)$. Theorem 1.26 comes to the rescue again, because if β *without* $\mathbf{v}$ is linearly independent and $\mathbf{v} \notin \text{span}(\beta)$, then $\gamma = \beta \cup \{\mathbf{v}\}$ must also be linearly independent. Our foolishness is exposed, for we now have a strictly larger extension of β which is also linearly independent, contradicting the maximality of β in this category. ■

[2] **Proof:** If $\mathbf{v}_i$ is in the span of $\{\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k\}$, say $\mathbf{v}_i = \sum_{j \neq i} \mathbf{v}_j$, then any vector $\mathbf{x}$ expressed as a linear combination $\mathbf{v} = \sum_{j=1}^k a_j \mathbf{v}_j$ can be reduced to a shorter linear combination of the remainder by substituting for $\mathbf{v}_i$ it's expression $\sum_{j \neq i} \mathbf{v}_j$ in terms of the other $\mathbf{v}_j$ and collecting terms.

Theorem 3.10 (Replacement) *Let $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $\gamma = \{\mathbf{w}_1, \ldots, \mathbf{w}_\ell\}$ be two nonempty subsets of $\mathbb{R}^n$, and suppose*

(1) β *is* linearly independent.

(2) γ *spans* $\mathbb{R}^n$.

Then, $k \leq \ell$, and as a result, we can break up γ into two disjoint pieces,

$$\begin{aligned} \gamma &= \gamma_1 \cup \gamma_2 \\ &= \{\mathbf{w}_1, \ldots, \mathbf{w}_k\} \cup \{\mathbf{w}_{k+1}, \ldots, \mathbf{w}_\ell\} \end{aligned}$$

then remove the first piece γ_1 and replace it with β—while still keeping the spanning:

$$\underbrace{\text{span}(\beta \cup \gamma_2)}_{\textit{we replaced } \gamma_1 \textit{ with } \beta} = \mathbb{R}^n = \text{span}(\gamma) = \text{span}(\gamma_1 \cup \gamma_2)$$

Example 3.11 Consider the set of four vectors

$$\gamma = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \left\{ \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 12 \\ 4 \end{pmatrix}, \begin{pmatrix} 70 \\ -22 \end{pmatrix} \right\}$$

in $\mathbb{R}^2$ from Example 1.2, and consider the linearly independent set consisting of the single vector

$$\beta = \{\mathbf{e}_1\} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$

Then we may relabel the $\mathbf{v}_i$, swapping $\mathbf{v}_1$ with $\mathbf{v}_2$, and breaking up γ into

$$\gamma = \gamma_1 \cup \gamma_2 = \{\mathbf{v}_1\} \cup \{\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \left\{ \begin{pmatrix} 3 \\ 0 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 12 \\ 4 \end{pmatrix}, \begin{pmatrix} 70 \\ -22 \end{pmatrix} \right\}$$

and with a little work we can see that

$$\boxed{\begin{aligned} \text{span}(\beta \cup \gamma_2) &= \text{span}(\mathbf{e}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4) \\ &= \text{span}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 12 \\ 4 \end{pmatrix}, \begin{pmatrix} 70 \\ -22 \end{pmatrix} \right) \\ &= \mathbb{R}^2 \\ &= \text{span}(\gamma) \end{aligned}}$$

For recall that, in order to show that any $\mathbf{b}$ in $\mathbb{R}^2$ lies in the span of $\beta \cup \gamma_2$, we must solve the system

$$\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{b} = a\mathbf{e}_1 + b\mathbf{v}_2 + c\mathbf{v}_3 + d\mathbf{v}_4 = \begin{pmatrix} 1 & -1 & 12 & 70 \\ 0 & 2 & 4 & -22 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

But since

$$\text{rref} \begin{pmatrix} 1 & -1 & 12 & 70 \\ 0 & 2 & 4 & -22 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 14 & 59 \\ 0 & 1 & 2 & -11 \end{pmatrix}$$

and the range of A is the range of rref(A), which is the span of its columns, we see that this is all of $\mathbb{R}^2$, so $\mathbf{b}$ can be gotten by row-reducing the augmented matrix $(A \mid \mathbf{b})$. ■

Exercise 3.12 Let

$$\gamma = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}, \begin{pmatrix} 11 \\ 0 \\ 9 \end{pmatrix} \right\}$$

and let

$$\beta = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \right\}$$

Show that γ spans $\mathbb{R}^3$, that β is linearly independent, and find a way to break up γ into $\gamma_1 \cup \gamma_2$ so that $\beta \cup \gamma_2$ spans $\mathbb{R}^3$. ■

Proof: This proof proceeds *by induction* on the size of β (see the marginal note for an explanation).

'This proof proceeds *by induction* on the size of β' means we prove it in two steps: Step 1, the **base step** or **base case**, is showing the truth of the theorem for β of size $k = 1$. Step 2 is the **inductive step**, which consists in *assuming* the truth of the theorem for β of *any* size $k \geq 1$, then *proving* it's truth somehow when β has the next size up, $k + 1$. When you combine the two steps you have everything, for all possible $k \geq 1$, for since you have $k = 1$ from the base step, you have $k = 2$ from the inductive step, and by iterating the inductive step you have $k = 3$, $k = 4$, etc. You have all possible k.

Step 1: *The base case,* $k = 1$. We suppose first that $\beta = \{\mathbf{v}_1\}$ consists of only one vector, $\mathbf{v}_1$. First of all, observe that $\mathbf{v}_1$ can't be the zero vector, for then we would have $1\mathbf{v}_1 = 1\mathbf{0} = \mathbf{0}$, a nontrivial representation of $\mathbf{0}$, whereas $\beta = \{\mathbf{v}_1\}$ is linearly independent and won't allow that. Second of all, $\mathbf{v}_1$ is a vector in $\mathbb{R}^n$, and by assumption $\mathbb{R}^n = \text{span}(\gamma)$, so $\mathbf{v}_1$ is a linear combination of the $\mathbf{w}_i$'s:

$$\mathbf{v}_1 = b_1\mathbf{w}_1 + \cdots + b_\ell\mathbf{w}_\ell$$

Combine these two observations, that $\mathbf{v}_1$ is nonzero in the span of the $\mathbf{w}_i$: the conclusion must be that *not all* the b_i can be zero! By relabeling[3] the $\mathbf{w}_i$'s appropriately we may assume that the first r of them, where $1 \leq r \leq k$, have nonzero coefficients, $b_i \neq 0$. Thus,

$$\mathbf{v}_1 = b_1\mathbf{w}_1 + \cdots + b_r\mathbf{w}_r$$

in terms of *nonzero* b_i's. Let's solve for $\mathbf{w}_1$, say:

$$\mathbf{w}_1 = \frac{1}{b_1}\mathbf{v}_1 - \frac{b_2}{b_1}\mathbf{w}_2 - \cdots - \frac{b_\ell}{b_1}\mathbf{w}_\ell$$

This is a happy thing, for it means we can *replace* $\mathbf{w}_1$ with $\mathbf{v}_1$ in γ and *still* get a spanning set: $\gamma = \gamma_1 \cup \gamma_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_\ell\}$ becomes $\gamma' = \beta \cup \gamma_2 = \{\mathbf{v}_1, \mathbf{w}_2, \ldots, \mathbf{w}_\ell\}$. Why is it still spanning, you say? Take any $\mathbf{x} \in \mathbb{R}^n$ and watch: since γ was spanning, we could write $\mathbf{x} = \sum_{j=1}^{k} a_j\mathbf{w}_j$, then substitute the above for $\mathbf{w}_1$:

$$\begin{aligned}
\mathbf{x} = \sum_{j=1}^{\ell} a_i\mathbf{w}_j &= a_1\mathbf{w}_1 + \sum_{j=2}^{\ell} a_j\mathbf{w}_j \\
&= a_1\left(\frac{1}{b_1}\mathbf{v}_1 - \sum_{j=2}^{\ell}\frac{b_j}{b_1}\mathbf{w}_j\right) + \sum_{j=2}^{\ell} a_j\mathbf{w}_j \\
&= \frac{a_1}{b_1}\mathbf{v}_1 + \sum_{j=2}^{\ell}\left(a_j - \frac{a_1 b_j}{b_1}\right)\mathbf{w}_j
\end{aligned}$$

[3] The reason we can relabel, we recall, is that β and γ here are considered *unordered* sets. Unordered means we can relabel as we please.

which exhibits $\mathbf{x}$ as an element of $\text{span}(\gamma')$. This completes the base case, $k = 1$.

We remark only that, for $k = 1$, we must have $1 \leq k \leq \ell$, *as a result* of Theorem 3.7, for the sizes of γ and γ' are both ℓ, and γ' is a spanning set, which can only be bigger than any linearly independent subset like β.

Step 2: *The inductive step.* We now assume the truth of the theorem whenever β has size $k \geq 1$, and show its truth for the case when β has size $k + 1$. For example, assuming the case of size 1 (which we can do, since we *proved* that case), this will show that it still works when β has size 2. Here we go: let

$$\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}\}$$

We're assuming this β is linearly independent, importantly, and also we're assuming that if we remove that last vector $\mathbf{v}_{k+1}$, then

$$\tilde{\beta} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$$

satisfies the theorem, i.e. we can replace $\gamma_1 = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ (after relabeling) with $\tilde{\beta}$ and *still* get a spanning set,

$$\text{span}(\tilde{\beta} \cup \gamma_2) = \text{span}\big(\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{w}_{k+1}, \ldots, \mathbf{w}_\ell\big) = \mathbb{R}^n$$

Let's cash in this last assumption: that vector we removed, $\mathbf{v}_{k+1}$, is a vector in $\mathbb{R}^n$, so it lies in $\text{span}(\tilde{\beta} \cup \gamma_2)$, meaning

$$\mathbf{v}_{k+1} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k + a_{k+1}\mathbf{w}_{k+1} + \cdots + a_\ell\mathbf{w}_\ell$$

Now, remember, β was linearly independent, so $\mathbf{v}_{k+1}$ can't be *merely* the first k terms with the rest of the $a_j = 0$, for that would contradict Theorem 1.26. And this shows, in particular, that $\ell > k$, and that not all a_j, for $k < j \leq \ell$, can be zero. By relabeling we can assume that $a_{k+1} \neq 0$, and use the above equation to solve for $\mathbf{w}_{k+1}$:

$$\begin{aligned}\mathbf{w}_{k+1} = -\Big(\frac{a_1}{a_{k+1}}\Big)\mathbf{v}_1 - \cdots - \Big(\frac{a_k}{a_{k+1}}\Big)\mathbf{v}_k \\ + \Big(\frac{1}{a_{k+1}}\Big)\mathbf{v}_{k+1} - \Big(\frac{a_{k+2}}{a_{k+1}}\Big)\mathbf{w}_{k+2} - \cdots - \Big(\frac{a_\ell}{a_{k+1}}\Big)\mathbf{w}_\ell\end{aligned}$$

This exhibits $\mathbf{w}_{k+1}$ as a linear combination of vectors in $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}\}$ and vectors in $\gamma_2 = \{\mathbf{w}_{k+2}, \ldots, \mathbf{w}_\ell\}$, and, as in the base case, this means we can replace the expression of any vector $\mathbf{v} \in \mathbb{R}^n = \text{span}(\tilde{\beta} \cup \gamma_2) = \mathbb{R}^n$ involving $\mathbf{w}_{k+1}$ with one involving $\mathbf{v}_{k+1}$ by substituting the above equation for $\mathbf{w}_{k+1}$. This shows that $\mathbb{R}^n = \text{span}(\beta \cup \tilde{\gamma}_2)$, where $\tilde{\gamma}_2 = \{\mathbf{w}_{k+2}, \ldots, \mathbf{w}_\ell\}$. ■

A direct consequence of this theorem is that any two bases have the same number of vectors:

Theorem 3.13 *Any two bases for $\mathbb{R}^n$ have the same number of vectors, namely n. This is the* ***dimension*** *of $\mathbb{R}^n$. This is true more generally for* any nontrivial subspace *V of $\mathbb{R}^n$: any two bases for V have the same number of vectors, $1 \leq k \leq n$, in which case we call k the* ***dimension*** *of V, and write*

$$\dim V = k, \quad \textit{where } k = |\beta| \textit{ is the size of any basis } \beta \textit{ for } V$$

Proof: We prove this only for the case of $\mathbb{R}^n$ for the moment. This proof is simple, and follows directly from the Replacement Theorem: use the linear independence of $\beta = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and the spanning property of $\gamma = \{\mathbf{c}_1, \ldots, \mathbf{c}_\ell\}$ to conclude that $k \leq \ell$. Then reverse their roles, use the linear independence of γ and the spanning property of β to conclude that $k \geq \ell$. Therefore $k = \ell$ and we are done. ■

Remark 3.14 The proof of the general statement, for *all* subspaces V of $\mathbb{R}^n$, is best approached from the abstract point of view, when we consider bases for the generic abstract vector space V. But one immediate result is we *do* have is the dimension of $\mathbb{R}^n$,

$$\dim \mathbb{R}^n = n$$

We know this because we *have* a basis for $\mathbb{R}^n$ already, the *standard basis*

$$\sigma = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$$

and this basis has n elements. ■

4 Linear Transformations and Bases

Recall the definition of a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, a function

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

satisfying

$$\begin{aligned} T(c\mathbf{x}) &= cT(\mathbf{x}) \\ T(\mathbf{x}+\mathbf{y}) &= T(\mathbf{x}) + T(\mathbf{y}) \end{aligned}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and all $c \in \mathbb{R}$. Matrix functions satisfy these conditions (Theorem 2.4, 'Matrices and Linear Transformations'). What's the *connection* between linear transformations and matrix functions? Answer: ***matrix functions are just linear transformations in standard coordinates***. Before we can explain this, let's nail down one *absolutely fundamental* general fact about linear transformations and bases: *the former are determined by the latter.*

Theorem 4.1 ***Any linear transformation*** $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ ***from*** $\mathbb{R}^n$ ***to*** $\mathbb{R}^m$ ***is*** **entirely determined** ***by it's action on any basis*** $\beta = \big(\mathbf{b}_1, \ldots, \mathbf{b}_n\big)$ ***for*** $\mathbb{R}^n$*:*

(1) *Given a linear transformation* $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$*,* all $\mathbf{y}$*-values of* T *in* $\mathbb{R}^m$ *are entirely determined by the handful of* $\mathbf{y}$*-values* T *takes on* β*,*

$$\mathbf{c}_1 = T(\mathbf{b}_1),\ \ldots,\ \mathbf{c}_n = T(\mathbf{b}_n)$$

(2) *If we define another function* S *on* β *by specifying the handful of* $\mathbf{y}$*-values to be these* $\mathbf{c}_j$*,*

$$S(\mathbf{b}_1) \overset{\text{def}}{=} \mathbf{c}_1,\ \ldots,\ S(\mathbf{b}_n) \overset{\text{def}}{=} \mathbf{c}_n$$

and then extend S *to all of* $\mathbb{R}^n$ *by linearity, meaning: if you expand any vector* $\mathbf{x} \in \mathbb{R}^n$ *in its unique* β*-coordinates,* $\mathbf{x} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{a}_n$*, and define*

$$\begin{aligned} S(\mathbf{x}) &= S(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{a}_n) \\ &\overset{\text{def}}{=} a_1 S(\mathbf{b}_1) + \cdots + a_n S(\mathbf{b}_n) \\ &= a_1\mathbf{c}_1 + \cdots + a_n\mathbf{c}_n \\ &\overset{\text{def}}{=} \mathbf{y} \end{aligned}$$

the result is

(a) S *is a linear transformation from* $\mathbb{R}^n$ *to* $\mathbb{R}^m$

(b) $S = T$

Remark 4.2 This theorem is of tremendous value, theoretically as well as practically. For instance, it allows us to *construct* linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ by *specifying* a $\mathbf{y}$-value for T on each basis vector $\mathbf{b}_i$ (with some purpose in mind, very likely), then *extending* T *by linearity*. We will frequently use this technique.

But the simplicity and cleanliness of the proof is really just a meditation on the meanings of the terms linear transformation and basis. Their interaction is simple and clean *by design*, because we selected the right features in each to make their interaction natural. Of course, we didn't do this arbitrarily, but used what we already knew about matrices and standard coordinates, so that our choices *naturally extended* the case of matrix functions in standard coordinates. ■

Proof: By Theorem 3.7 we know that any vector $\mathbf{x}$ in $\mathbb{R}^n$ has a *unique* β-coordinate representation, $\mathbf{x} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n$. To prove (1) it suffices only to observe that $\mathbf{y} = a_1\mathbf{c}_1 + \cdots + a_n\mathbf{c}_n$ is entirely determined by the β-coordinates of $\mathbf{x}$, the coefficients a_j, and the fact that T's linearity allows the equality

$$T(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n) = a_1T(\mathbf{b}_1) + \cdots + a_nT(\mathbf{b}_n)$$

Thus, to clinch this we must show (2), that *if* $S : \mathbb{R}^n \to \mathbb{R}^m$ is defined on β and extended to $\mathbb{R}^n$ *in the same way as* T,

$$S(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n) = a_1S(\mathbf{b}_1) + \cdots + a_nS(\mathbf{b}_n)$$

then S is linear—$S(\mathbf{x}+\mathbf{y}) = S(\mathbf{x}) + S(\mathbf{y})$ and $S(c\mathbf{x}) = cS(\mathbf{x})$—and in this case must equal T, which also satisfies this identity. Let's start with proving $S(\mathbf{x}) + S(\mathbf{y})$: first, expand each in β:

$$\begin{aligned}\mathbf{x} &= a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n \\ \mathbf{y} &= b_1\mathbf{b}_1 + \cdots + b_n\mathbf{b}_n\end{aligned}$$

and compute,

$$\begin{aligned}S(\mathbf{x}+\mathbf{y}) &= S\Big(\big(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n\big) + \big(b_1\mathbf{b}_1 + \cdots + b_n\mathbf{b}_n\big)\Big) \\ &= S\Big((a_1+b_1)\mathbf{b}_1 + \cdots + (a_n+b_n)\mathbf{b}_n\Big) \qquad (\textit{associativity}) \\ &= (a_1+b_1)S(\mathbf{b}_1) + \cdots + (a_n+b_n)S(\mathbf{b}_n) \qquad (\textit{definition of } S) \\ &= \Big[a_1S(\mathbf{b}_1) + \cdots + a_nS(\mathbf{b}_n)\Big] + \Big[b_1S(\mathbf{b}_1) + \cdots + b_nS(\mathbf{b}_n)\Big] \qquad (\textit{distributivity}) \\ &= S\big(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n\big) + S\big(b_1\mathbf{b}_1 + \cdots + b_n\mathbf{b}_n\big) \qquad (\textit{definition of } S) \\ &= S(\mathbf{x}) + S(\mathbf{y})\end{aligned}$$

You might say this proof is tedious and repetitious. It is, in a way. But in another way, it is a **recitation** of the ideas. Recitation is like practicing scales in music, or lifting weights at the gym. You don't do it for its own sake, but in order to develop certain skills and certain muscles. Proofs are recitations in this way. Use them to **review** the ideas, and learn their interconnections.

Similarly,

$$\begin{aligned}S(c\mathbf{x}) &= S\Big(c\big(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n\big)\Big) \\ &= S\Big((ca_1)\mathbf{b}_1 + \cdots + (ca_n)\mathbf{b}_n\Big) \qquad (\textit{distributivity \& associativity}) \\ &= (ca_1)S(\mathbf{b}_1) + \cdots + (ca_n)S(\mathbf{b}_n) \qquad (\textit{definition of } S) \\ &= c\Big[a_1S(\mathbf{b}_1) + \cdots + a_nS(\mathbf{b}_n)\Big] \qquad (\textit{associativity \& distributivity}) \\ &= cS\Big(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n\Big) \qquad (\textit{definition of } S) \\ &= cS(\mathbf{x})\end{aligned}$$

Finally, since S agrees with T on all basis vectors and hence on all $\mathbf{x} \in \mathbb{R}^n$ by linearity, $S = T$ ■

5 Invertible Matrices and Bases

Notation 5.1 *In what follows, I'll be using the following equivalent notation for the range of a matrix* $A \in M_{m,n}(\mathbb{R})$,

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

It just emphasizes the $A\mathbf{x}$ *part of* $\mathbf{y} = A\mathbf{x}$. *In what follows, we also use this notation on a finite set of vectors*

$$\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$$

We write, accordingly,

$$A\beta = \{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\}$$

■

This notation, too, is very standard in the literature. Both are used equally.

Theorem 5.2 (The Action of an Invertible Matrix) *Let* $A \in \mathrm{GL}(n, \mathbb{R})$ *be an invertible square matrix. If* $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ *is a collection of vectors in* $\mathbb{R}^n$*, then A's action on* β *is governed by the following rules:*

(1) *If* $\mathrm{span}(\beta) = \mathbb{R}^n$*, then* $\mathrm{span}(A\beta) = \mathbb{R}^n$.

(2) β *is linearly independent iff* $A\beta$ *is linearly independent.*

(3) β *is a basis for* $\mathbb{R}^n$ *iff* $A\beta$ *is a basis for* $\mathbb{R}^n$.

This is the key part. The other two points are preliminary.

Proof: (1) Suppose that $\mathrm{span}(\beta) = \mathbb{R}^n$. Since A is invertible, it is onto and one-to-one. Use the onto part, which says $R(A) = A(\mathbb{R}^n) = \mathbb{R}^n$, and combine it with the *linearity* of A to get

$$\begin{aligned}
\mathbb{R}^n &= A(\mathbb{R}^n) \\
&= A(\mathrm{span}(\beta)) && \textit{(our assumption)} \\
&= \left\{ A\Big(\sum_{i=1}^k a_i\mathbf{v}_i\Big) \;\middle|\; a_i \in \mathbb{R} \right\} && \textit{(definition of span)} \\
&= \left\{ \sum_{i=1}^k a_i A\mathbf{v}_i \;\middle|\; a_i \in \mathbb{R} \right\} && \textit{(linearity of } A\textit{)} \\
&= \mathrm{span}(A\beta) && \textit{(definition of span)}
\end{aligned}$$

(2) Suppose β is linearly independent, and consider a possible representation of $\mathbf{0}$ in $A\beta = \{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\}$,

$$a_1A\mathbf{v}_1 + \cdots + a_kA\mathbf{v}_k = \mathbf{0} \tag{5.1}$$

Now use the *one-to-one* aspect of A in combination with its *linearity*:

$$\mathbf{0} = a_1A\mathbf{v}_1 + \cdots + a_kA\mathbf{v}_k = A\big(a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k\big)$$

and notice how this exhibits the vector $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$ as an element of the null space of A. Since A is one-to-one, it's null space is trivial (Proposition 3.9), $N(A) = \{\mathbf{0}\}$. We conclude that

$$a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0} \tag{5.2}$$

This argument proceeds in two moves: (i) the linearity of A allows passage between $A(\sum_{i=1}^k a_i\mathbf{v}_i)$ and $\sum_{i=1}^k a_iA\mathbf{v}_i$, i.e. between $\mathrm{span}(\beta)$ and $\mathrm{span}(A\beta)$. (ii) The 1-1 assumption on A then allows (by Prop 3.9, *'Matrices and Linear Transformations'*) passage from $N(A) = \{\mathbf{0}\}$ and the linear independence of β and $A\beta$, which is about $\mathbf{0}$'s representation being trivial.

Now use the linear independence of β to conclude that in the representation (5.2) above all the coefficients a_i are 0. But these are the same a_i as in our starting representation of $\mathbf{0}$, (5.1). We conclude that $A\beta$ is linearly independent.

If $A\beta$ is linearly independent and A is invertible, then we can we conclude β is linearly directly, as we did above, or we can be clever and use what we've already developed. Since A is invertible, so is A^{-1} (Theorem 4.6, '*Matrices and Linear Transformations*'), so apply A^{-1} to $A\beta$ to conclude that $\beta = I\beta = (A^{-1}A)\beta = A^{-1}(A\beta)$ is linearly independent.

(3) This is just a combination of points (1) and (2): Since A is invertible, β is spanning iff $A\beta$ is, and β is linearly independent iff $A\beta$ is. Thus, β is a basis iff $A\beta$ is. ■

Remark 5.3 What we know so far: **If** we know that *A is invertible*, **then** we know that *A takes bases β to bases $A\beta$*. This is what the previous theorem says. We also have an earlier theorem about invertible matrices (the combination of Corollary 3.8 and Proposition 3.9, in '*Matrices and Linear Transformations*'), that *A is invertible* **iff** *its columns form a basis $\alpha = \big(\mathbf{a}_1, \dots, \mathbf{a}_n\big)$*. Can we combine these two statements into one general one? Answer: Yes, but we need to consider change-of-coordinates transformations and their matrix representations, which is in the next section. In the meantime, let us establish a converse to the above statement in red, so that we get a *logical equivalence* between *invertibility* and the *carrying of bases to bases*: ■

Theorem 5.4 (Invertible Matrices Preserve Bases) *Let $A \in M_n(\mathbb{R})$ be a square matrix and $\beta = \big(\mathbf{b}_1, \dots, \mathbf{b}_n\big)$ a basis for $\mathbb{R}^n$. Then $A \in \mathrm{GL}(n, \mathbb{R})$ iff $A\beta$ is a basis for $\mathbb{R}^n$.*

Proof: We have already shown that *if* A is invertible, *then* $A\beta$ is a basis whenever β is. Now suppose $A\beta$ is a basis whenever β is. To show that A is invertible, we show that it is one-to-one and onto. Starting with one-to-one, we use Proposition 3.9, '*Matrices and Linear Transformations*', and show the logically equivalent thing, that $N(A) = \{\mathbf{0}\}$: let $\mathbf{x} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n$ be the unique β-coordinate representation of a null vector $\mathbf{x} \in N(A)$ and let us prove that $\mathbf{x} = 0$ by proving that all $a_j = 0$. Well, $A\beta$ is a basis if β is, by our assumption, so it is linearly independent—use this and the linearity of A:

$$\begin{aligned} \mathbf{0} &= A\mathbf{x} \qquad \big(\textit{because } \mathbf{x} \in N(A)\big) \\ &= A\big(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n\big) \\ &= a_1\big(A\mathbf{b}_1\big) + \cdots + a_n\big(A\mathbf{b}_n\big) \qquad \big(A \textit{ is linear}\big) \end{aligned}$$

but since $A\beta = \big(A\mathbf{b}_1, \dots, A\mathbf{b}_n\big)$ is a basis, it is linearly independent, which forces all a_j to be 0. Thus, $a_1 = \cdots = a_n = 0$, and so $\mathbf{x} = \sum_{j=1}^n a_j\mathbf{b}_j = \mathbf{0}$ (see Observation 3.4 above). Thus, $N(A) = \{\mathbf{0}\}$ and A is one-to-one.

To show that A is onto, choose any $\mathbf{y} \in \mathbb{R}^n$ and let us find an $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{y}$. Use the basis $A\beta$ to expand $\mathbf{y}$ uniquely in $A\beta$-coordinates, plus the linearity of A:

$$\begin{aligned} \mathbf{y} &= a_1\big(A\mathbf{b}_1\big) + \cdots + a_n\big(A\mathbf{b}_n\big) \\ &= A\big(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n\big) \end{aligned}$$

Thus, if we define $\mathbf{x}$ to be $a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n$, then we have shown $A\mathbf{x} = \mathbf{y}$. ■

6 A Deeper Study of Coordinates in $\mathbb{R}^n$

Definition 6.1 (Coordinate Representations of Vectors) Let $\beta = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ and $\gamma = (\mathbf{c}_1, \ldots, \mathbf{c}_n)$ be any two bases for $\mathbb{R}^n$, and let $\sigma = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be the standard basis. The **β-coordinates** of any vector $\mathbf{x} \in \mathbb{R}^n$, we recall, are the coefficients in the unique expression

$$\mathbf{x} = b_1\mathbf{b}_1 + \cdots + b_n\mathbf{b}_n$$

while the **γ-coordinates** of $\mathbf{x}$ are the coefficients in the unique expression

$$\mathbf{x} = c_1\mathbf{c}_1 + \cdots + c_n\mathbf{c}_n$$

These may be considered two *representations* of $\mathbf{x}$, though we typically reserve the term **representation** (or **coordinate vector**) for the *column vectors* containing the coordinates: the β- and γ-representations are denoted, respectively,

$$[\mathbf{x}]_\beta = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \qquad [\mathbf{x}]_\gamma = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

The **standard representation** is the σ-coordinate representation:

$$[\mathbf{x}]_\sigma = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{x}$$

which of course follows from the fact that $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$. ■

Definition 6.2 In order to transition between two coordinate representations $[\mathbf{x}]_\beta$ and $[\mathbf{x}]_\gamma$ we emply the **change-of-coordinates matrix**

$$M_{\beta,\gamma} = \begin{pmatrix} | & & | \\ [\mathbf{b}_1]_\gamma & \cdots & [\mathbf{b}_n]_\gamma \\ | & & | \end{pmatrix}$$

which changes β-coordinates to γ-coordinates (a theorem to be proved below), as follows:

$$\boxed{[\mathbf{x}]_\gamma = M_{\beta,\gamma}[\mathbf{x}]_\beta}$$ ■

Definition 6.3 (Coordinate Representations of Matrices) We can also represent matrices $A \in M_{m,n}(\mathbb{R})$ in other coordinate systems β for $\mathbb{R}^n$ and γ for $\mathbb{R}^m$, and do so *in a way that preserves products* and *linearity*: let

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

and let $\beta = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ be a basis for $\mathbb{R}^n$ and $\gamma = (\mathbf{c}_1, \dots, \mathbf{c}_m)$ be a basis for $\mathbb{R}^m$. We can represent each product-vector $A\mathbf{b}_i$ in γ-coordinates,

$$[A\mathbf{b}_i]_\gamma$$

and then arrange these in columns of the (β, γ)**-representation** matrix of A,

$$[A]_{\beta,\gamma} \stackrel{\text{def}}{=} \begin{pmatrix} | & & | \\ [A\mathbf{b}_1]_\gamma & \cdots & [A\mathbf{b}_n]_\gamma \\ | & & | \end{pmatrix}$$

In the special case that $n = m$, so that $A \in M_n(\mathbb{R})$ is square, and $\beta = \gamma$, we write

$$[A]_\beta \ \textbf{instead of} \ [A]_{\beta,\beta}$$ ■

We proceed similarly with abstract, coordinate-free, linear transformations: we coordinatize them with respect to two bases, β for the domain, $\mathbb{R}^n$, and γ with respect to $\mathbb{R}^m$.

Definition 6.4 (Coordinate Representations of Linear Transformations) Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and choose bases $\beta = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ for $\mathbb{R}^n$ and $\gamma = (\mathbf{c}_1, \dots, \mathbf{c}_m)$ for $\mathbb{R}^m$. Define the (β, γ)**-representation** of T,

$$[T]_{\beta,\gamma} \stackrel{\text{def}}{=} \begin{pmatrix} | & & | \\ [T(\mathbf{b}_1)]_\gamma & \cdots & [T(\mathbf{b}_n)]_\gamma \\ | & & | \end{pmatrix}$$

In the special case that $n = m$, so that $T \in \mathcal{L}(\mathbb{R}^n)$, and $\beta = \gamma$, we write

$$[T]_\beta \ \textbf{instead of} \ [T]_{\beta,\beta}$$ ■

Remark 6.5 These two definitions agree, seeing as matrix functions are a species of linear transformation. In fact, if we consider a matrix $A \in M_{m,n}(\mathbb{R})$ as an abstract linear transformation (coordinate-free), let us denote it

$$T_A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), \quad T_A(\mathbf{x}) \overset{\text{def}}{=} A\mathbf{x}$$

then T_A's particular numerical realization in terms of A's entries a_{ij} is transparently the *standard representation* of T_A:

$$\begin{aligned} [T_A]_{\sigma_n,\sigma_m} &= \begin{pmatrix} [T_A(\mathbf{e}_1)]_{\sigma_m} & \cdots & [T_A(\mathbf{e}_n)]_{\sigma_m} \end{pmatrix} \\ &= \begin{pmatrix} [A\mathbf{e}_1]_{\sigma_m} & \cdots & [A\mathbf{e}_n]_{\sigma_m} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix} \\ &= A \end{aligned}$$

Conversely, every abstract (coordinate-free) linear transformation has a **standard matrix representation**,

$$A \overset{\text{def}}{=} [T]_{\sigma_n,\sigma_m}$$

In this case, we have

$$[T_A]_{\sigma_n,\sigma_m} = A = [T]_{\sigma_n,\sigma_m}$$

as we will see below, on account of $[T]_{\beta,\gamma}[\mathbf{x}]_\beta = [T(\mathbf{x})]_\gamma$. Thus, the connection between the abstract (coordinate-free) linear transformation and its standard matrix A is precisely the standard coordinate representation of T. We can go from T to A via this representation, and we can go from A to T_A and arrive back at T, which *is* T_A! ■

Exercise 6.6 Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Prove that T is one-to-one iff its null space

$$N(T) \overset{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^n \mid T(\mathbf{x}) = \mathbf{0}\}$$

is trivial. The proof is virtually that of Proposition 3.9, *'Matrices, and Linear Transformations'*. ■

Let us pause and put some of this togther into a sentence. The easiest way is *by means* of linear transformations, since those *are* coordinate-free to begin with. Here's the way:

Definition 6.7 The coordinate representation $[\mathbf{x}]_\beta$ may be understood more *dynamically*, as the *result* of an operation. The operation is a *function* whose definition *requires* the basis β,

$$\varphi_\beta : \mathbb{R}^n \to \mathbb{R}^n$$

It is called the β-**coordinate map**, and is defined by

$$\varphi_\beta(\mathbf{x}) \overset{\text{def}}{=} [\mathbf{x}]_\beta$$ ■

Proposition 6.8 *Each coordinate map φ_β is an invertible linear transformation, and satisfies*

$$\boxed{\varphi_\beta(\mathbf{b}_j) = \mathbf{e}_j}$$

for each $\mathbf{b}_j \in \beta$.

Proof: Since each $\mathbf{b}_j$ has β-representation

$$\mathbf{b}_j = 0\mathbf{b}_1 + \cdots + 1\mathbf{b}_j + \cdots + 0\mathbf{b}_n$$

its β-coordinate vector is

$$\varphi_\beta(\mathbf{b}_j) = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow j\text{th slot} \;\; = \;\; \mathbf{e}_j$$

because the 1 is in the jth slot. From this observation it is easy to see that φ_β is linear, since if $\mathbf{x} = \sum_{j=1}^n b_j\mathbf{b}_j$ and $\mathbf{y} = \sum_{j=1}^n c_j\mathbf{b}_j$ are two vectors in $\mathbb{R}^n$, then

$$\begin{aligned}
\varphi_\beta(\mathbf{x}+\mathbf{y}) &= \varphi_\beta\Big(\sum_{j=1}^n b_j\mathbf{b}_j + \sum_{j=1}^n c_j\mathbf{b}_j\Big) \\
&= \varphi_\beta\Big(\sum_{j=1}^n (b_j + c_j)\mathbf{b}_j\Big) \\
&= \begin{pmatrix} b_1 + c_1 \\ \vdots \\ b_n + c_n \end{pmatrix} \\
&= \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} + \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \\
&= \varphi_\beta\Big(\sum_{j=1}^n b_j\mathbf{b}_j\Big) + \varphi_\beta\Big(\sum_{j=1}^n c_j\mathbf{b}_j\Big) \\
&= \varphi_\beta(\mathbf{x}) + \varphi_\beta(\mathbf{y})
\end{aligned}$$

and similarly if $c \in \mathbb{R}$ and $\mathbf{x} = \sum_{j=1}^n b_j\mathbf{b}_j \in \mathbb{R}^n$,

$$\begin{aligned}
\varphi_\beta(c\mathbf{x}) &= \varphi_\beta\Big(c\sum_{j=1}^n b_j\mathbf{b}_j\Big) \\
&= \varphi_\beta\Big(\sum_{j=1}^n cb_j\mathbf{b}_j\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \begin{pmatrix} cb_1 \\ \vdots \\ cb_n \end{pmatrix} \\
&= c \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \\
&= c\varphi_\beta\Big(\sum_{j=1}^{n} b_j \mathbf{b}_j\Big) \\
&= c\varphi_\beta(\mathbf{x})
\end{aligned}
$$

OK, $\varphi_\beta \in \mathcal{L}(\mathbb{R}^n)$. Let's now show it's one-to-one, and afterwards that it's onto: We show $N(\varphi_\beta)$ is trivial and refer to Exercise 6.6 for a proof that this is equivalent to T being one-to-one. Let $\mathbf{x} = \sum_{j=1}^{n} b_j \mathbf{b}_j \in N(\varphi_\beta)$, then

$$
\mathbf{0} = \varphi_\beta(\mathbf{x}) = \varphi_\beta\Big(\sum_{j=1}^{n} b_j \mathbf{b}_j\Big) = \sum_{j=1}^{n} b_j \varphi_\beta(\mathbf{b}_j) = \sum_{j=1}^{n} b_j \mathbf{e}_j = [\mathbf{x}]_\beta
$$

which means all $b_i = 0$, and therefore $\mathbf{x} = \mathbf{0}$. Let us now show that φ_β is onto: choose an arbitrary vector $\mathbf{y} = \sum_{j=1} y_j \mathbf{e}_j \in \mathbb{R}^n$ and observe that it is in φ_β's range:

$$
\mathbf{y} = \sum_{j=1}^{n} y_j \mathbf{e}_j = \sum_{j=1}^{n} y_j \varphi_\beta(\mathbf{b}_j) = \varphi_\beta\Big(\sum_{j=1}^{n} y_j \mathbf{b}_j\Big) = \varphi_\beta(\mathbf{x})
$$

where $\mathbf{x} = \sum_{j=1}^{n} y_j \mathbf{b}_j \in \mathbb{R}^n$. Thus, φ_β is linear, one-to-one, and onto. ■

Let us try to understand this, now. Any invertible linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ is called an **isomorphism** of $\mathbb{R}^n$ (with itself). We distinguish between isomorphisms and invertible matrices just in the same way that we distinguish between linear transformations and matrices. One is coordinate-free, the other is not—it is the *standard* representation of the former.

Notation 6.9 *Let us remind ourselves now of the notation for the* ***set of all isomorphisms*** *between $\mathbb{R}^n$ and itself,*

$$\mathrm{GL}(\mathbb{R}^n) \stackrel{\text{def}}{=} \{T \in \mathcal{L}(\mathbb{R}^n) \mid T \text{ is invertible}\}$$

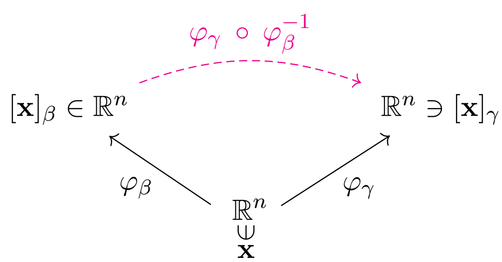
Like $\mathrm{GL}(n, \mathbb{R})$, it is a group. It is a more abstract group, because it is defined solely in terms of properties. Yet it is still rooted in real numbers, which are present in the background, in the appearance of $\mathbb{R}^n$. Here is a list of the coordinatized and coordinate-free notations developed so far:

Coordinatized	*Coordinate-Free*
$M_{m,n}(\mathbb{R})$	$\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$
$M_n(\mathbb{R})$	$\mathcal{L}(\mathbb{R}^n)$
$\mathrm{GL}(n, \mathbb{R})$	$\mathrm{GL}(\mathbb{R}^n)$

Let us now try to understand their interrelations. ■

Here is a diagrammatic picture of coordinate changes on vectors:

This diagram compresses the essential idea behind coordinate changes. Coordinates are a plaything, we **manipulate** coordinates for our own benefit, and this manipulation happens **via** coordinate maps. *We* are masters of our destiny if we can make coordinate maps our constant companions.



The composition of φ_γ with φ_β^{-1} is a linear transformation, and the composition of two isomorphisms is an isomorphism, according to the following exercises:

Exercise 6.10 Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and let $S \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$. Show that the composition $S \circ T : \mathbb{R}^n \to \mathbb{R}^m$ is also linear, and so belongs to $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$. ■

Exercise 6.11 Let $S, T \in \mathrm{GL}(\mathbb{R}^n)$ be linear isomorphisms (invertible linear operators). Show that S^{-1} and $S \circ T$ are also isomorphisms, so that they belong to $\mathrm{GL}(\mathbb{R}^n)$, and that

$$\begin{aligned}(S^{-1})^{-1} &= S \\ (S \circ T)^{-1} &= T^{-1} \circ S^{-1}\end{aligned}$$

This shows that $\mathrm{GL}(\mathbb{R}^n)$ is a group, just like $\mathrm{GL}(n, \mathbb{R})$. [*Hint: the proofs are in fact very similar to those of* $\mathrm{GL}(n, \mathbb{R})$*!*] ■

Proposition 6.12 *The* ***standard representation*** *of the composition*

$$\varphi_\gamma \circ \varphi_\beta^{-1} \in \mathrm{GL}(\mathbb{R}^n)$$

is just the change-of-coordinates matrix $M_{\beta,\gamma}$*:*

$$[\varphi_\gamma \circ \varphi_\beta^{-1}]_\sigma = M_{\beta,\gamma}$$

and the standard representation of its inverse

$$(\varphi_\gamma \circ \varphi_\beta^{-1})^{-1} = \varphi_\beta \circ \varphi_\gamma^{-1}$$

is the change-of-coordinates matrix $M_{\gamma,\beta}$*, which is the inverse of* $M_{\beta,\gamma}$*:*

$$M_{\gamma,\beta} = [\varphi_\beta \circ \varphi_\gamma^{-1}]_\sigma = [(\varphi_\gamma \circ \varphi_\beta^{-1})^{-1}]_\sigma = [(\varphi_\gamma \circ \varphi_\beta^{-1})]_\sigma^{-1} = M_{\beta,\gamma}^{-1}$$

Proof: The standard representation of any linear operator $T \in \mathcal{L}(\mathbb{R}^n)$ is

$$[T]_\sigma = \big([T(\mathbf{e}_1)]_\sigma \quad \cdots \quad [T(\mathbf{e}_n)]_\sigma\big)$$

So, too, in the case of $\varphi_\gamma \circ \varphi_\beta^{-1}$. But let us recall Proposition 6.8, which says that $\varphi_\beta(\mathbf{b}_j) = \mathbf{e}_j$. For then, $\mathbf{b}_j = \varphi_\beta^{-1}(\mathbf{e}_j)$, and

$$\begin{aligned}[\varphi_\gamma \circ \varphi_\beta^{-1}]_\sigma &= \Big([(\varphi_\gamma \circ \varphi_\beta^{-1})(\mathbf{e}_1)]_\sigma \quad \cdots \quad [(\varphi_\gamma \circ \varphi_\beta^{-1})(\mathbf{e}_n)]_\sigma\Big) \\ &= \Big([\varphi_\gamma(\mathbf{b}_1)]_\sigma \quad \cdots \quad [\varphi_\gamma(\mathbf{b}_n)]_\sigma\Big) \\ &= \Big(\varphi_\gamma(\mathbf{b}_1) \quad \cdots \quad \varphi_\gamma(\mathbf{b}_n)\Big) \\ &= \Big([\mathbf{b}_1]_\gamma \quad \cdots \quad [\mathbf{b}_n]_\gamma\Big) \\ &= M_{\beta,\gamma}\end{aligned}$$

The other direction is similar, and is left as an exercise. ■

Theorem 6.13 *Let $\mathbf{x} \in \mathbb{R}^n$ and consider two bases $\beta = \big(\mathbf{b}_1, \ldots, \mathbf{b}_n\big)$ and $\gamma = \big(\mathbf{c}_1, \ldots, \mathbf{c}_n\big)$ for $\mathbb{R}^n$. The transformation rule for changing coordinate representations of $\mathbf{x}$ from β-coordinates to γ-coordinates is*

$$\boxed{[\mathbf{x}]_\gamma = M_{\beta,\gamma}[\mathbf{x}]_\beta}$$

Proof: If you understand the standard representation of $\varphi_\gamma \circ \varphi_\beta^{-1}$ as just this function applied to the identity matrix, whose columns are the standard basis vectors $\mathbf{e}_j$,

$$\begin{aligned} M_{\beta,\gamma} &= [\varphi_\gamma \circ \varphi_\beta^{-1}]_\sigma \\ &= \Big([(\varphi_\gamma \circ \varphi_\beta^{-1})(\mathbf{e}_1)]_\sigma \quad \cdots \quad [(\varphi_\gamma \circ \varphi_\beta^{-1})(\mathbf{e}_n)]_\sigma\Big) \\ &= (\varphi_\gamma \circ \varphi_\beta^{-1}) \circ \big(\mathbf{e}_1 \quad \cdots \quad \mathbf{e}_n\big) \\ &= \varphi_\gamma \circ \varphi_\beta^{-1} \circ I_n \end{aligned}$$

then it's easy: since $\varphi_\beta^{-1} \circ I_n \circ \varphi_\beta = \varphi_\beta^{-1} \circ \varphi_\beta = I_n$, we have

$$M_{\beta,\gamma}[\mathbf{x}]_\beta = \big(\varphi_\gamma \circ \varphi_\beta^{-1} \circ I_n \circ \varphi_\beta\big)(\mathbf{x}) = (\varphi_\gamma \circ I_n)(\mathbf{x}) = \varphi_\gamma(\mathbf{x}) = [\mathbf{x}]_\gamma \qquad \blacksquare$$

Corollary 6.14 *Let $\beta = \big(\mathbf{b}_1, \ldots, \mathbf{b}_n\big)$ be a basis for $\mathbb{R}^n$. Then*

$$\boxed{M_{\beta,\sigma} = \begin{pmatrix} | & & | \\ \mathbf{b}_1 & \cdots & \mathbf{b}_n \\ | & & | \end{pmatrix}}$$

so that for any $\mathbf{x} \in \mathbb{R}^n$ we can find $[\mathbf{x}]_\beta$ by inverting this matrix,

$$\boxed{[\mathbf{x}]_\beta = M_{\sigma,\beta}[\mathbf{x}]_\sigma = M_{\beta,\sigma}^{-1}\mathbf{x}} \qquad \blacksquare$$

The computational problem of finding $[\mathbf{x}]_\beta$ resolves into a matrix inversion and its product with $\mathbf{x}$. What about switching between different representations? In order to appreciate the idea behind that rule, it is necessary to return to representations of linear transformations in general, and their transformation rules. First, an observation:

Observation 6.15 In terms of coordinate maps,

$$[T]_{\beta,\gamma} = \Big([T(\mathbf{b}_1)]_\gamma \quad \cdots \quad [T(\mathbf{b}_n)]_\gamma\Big) = \Big((\varphi_\gamma \circ T \circ \varphi_\beta^{-1})(\mathbf{e}_1) \quad \cdots \quad (\varphi_\gamma \circ T \circ \varphi_\beta^{-1})(\mathbf{e}_n)\Big)$$

i.e.

$$\boxed{[T]_{\beta,\gamma} = [\varphi_\gamma \circ T \circ \varphi_\beta^{-1}]_\sigma}$$

Indeed, we can even understand the change-of-coordinates operator $\varphi_\gamma \circ \varphi_\beta^{-1}$ like this, as ***representing the identity operator*** $I \in \mathrm{GL}(\mathbb{R}^n)$,

$$\boxed{M_{\beta,\gamma} = [\varphi_\gamma \circ \varphi_\beta^{-1}]_\sigma = [\varphi_\gamma \circ I \circ \varphi_\beta^{-1}]_\sigma = [I]_{\beta,\gamma}} \qquad \blacksquare$$

Now consider the following generalization of Corollary 6.14

Theorem 6.16 *Consider bases $\beta = \big(\mathbf{b}_1, \dots, \mathbf{b}_n\big)$ for $\mathbb{R}^n$ and $\gamma = \big(\mathbf{c}_1, \dots, \mathbf{c}_m\big)$ for $\mathbb{R}^m$, and let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ be a linear transformation. Then, for all $\mathbf{v} \in \mathbb{R}^n$ we have*

$$\boxed{[T(\mathbf{v})]_\gamma = [T]_{\beta,\gamma}[\mathbf{v}]_\beta}$$

which may neatly be understood in the following diagram:

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{\ T\ } & \mathbb{R}^m \\ \varphi_\beta \big\downarrow & & \big\downarrow \varphi_\gamma \\ \mathbb{R}^n & \xrightarrow[T_{[T]_{\beta,\gamma}}]{} & \mathbb{R}^n \end{array}, \qquad \begin{array}{c} T_{[T]_{\beta,\gamma}} \circ \varphi_\beta = \varphi_\gamma \circ T \\ \text{or} \\ T_{[T]_{\beta,\gamma}} = \varphi_\gamma \circ T \circ \varphi_\beta^{-1} \end{array}$$

Proof: Expand $\mathbf{v}$ in β-coordinates, $\mathbf{v} = b_1\mathbf{b}_1 + \cdots + b_n\mathbf{b}_n$, and use the linearity of T:

$$T(\mathbf{v}) = T\Big(\sum_{j=1}^{n} b_j\mathbf{b}_j\Big) = \sum_{j=1}^{n} b_j T(\mathbf{b}_j)$$

then take the γ-representation φ_γ of both sides, using the linearity of φ_γ on the RHS,

$$\begin{aligned} [T(\mathbf{v})]_\gamma &= \varphi_\gamma\big(T(\mathbf{v})\big) \\ &= \varphi_\gamma\Big(\sum_{j=1}^{n} b_j T(\mathbf{b}_j)\Big) \\ &= \sum_{j=1}^{n} b_j \varphi_\gamma\big(T(\mathbf{b}_j)\big) \\ &= \sum_{j=1}^{n} b_j \big[T(\mathbf{b}_j)\big]_\gamma \\ &= \Big(\big[T(\mathbf{b}_1)\big]_\gamma \quad \cdots \quad \big[T(\mathbf{b}_n)\big]_\gamma\Big) \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \\ &= [T]_{\beta,\gamma}[\mathbf{v}]_\beta \end{aligned}$$

We are basically done. Proposition 6.8 gives the last detail: since $\varphi_\beta(\mathbf{b}_j) = \mathbf{e}_j$, we can write $[T(\mathbf{v})]_\gamma = (\varphi_\gamma \circ T \circ \varphi_\beta^{-1})(\mathbf{v}) = \sum_{j=1}^{n} b_j(\varphi_\gamma \circ T \circ \varphi_\beta^{-1})(\mathbf{e}_j)$, which is just the statement $T_{[T]_{\beta,\gamma}} = \varphi_\gamma \circ T \circ \varphi_\beta^{-1}$. ■

Theorem 6.17 *Consider three bases for $\mathbb{R}^n$, $\mathbb{R}^m$ and $\mathbb{R}^p$, respectively*

$$\beta = \big(\mathbf{b}_1, \dots, \mathbf{b}_n\big), \quad \gamma = \big(\mathbf{c}_1, \dots, \mathbf{c}_m\big), \quad \delta = \big(\mathbf{d}_1, \dots, \mathbf{d}_p\big)$$

and consider any two linear transformations $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $S \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$. Then

$$\boxed{[S \circ T]_{\beta,\delta} = [S]_{\gamma,\delta}[T]_{\beta,\gamma}}$$

Proof: By the previous theorem we have

$$\begin{aligned}
[S \circ T]_{\beta,\delta} &= \Big([(S\circ T)(\mathbf{b}_1)]_\delta \ \cdots \ [(S\circ T)(\mathbf{b}_n)]_\delta \Big) \\
&= \Big([S(T(\mathbf{b}_1))]_\delta \ \cdots \ [S(T(\mathbf{b}_n))]_\delta \Big) \\
&= \Big([S]_{\gamma,\delta}[T(\mathbf{b}_1)]_\gamma \ \cdots \ [S]_{\gamma,\delta}[T(\mathbf{b}_n)]_\gamma \Big) \\
&= [S]_{\gamma,\delta}\Big([T(\mathbf{b}_1)]_\gamma \ \cdots \ [T(\mathbf{b}_n)]_\gamma \Big) \\
&= [S]_{\gamma,\delta}[T]_{\beta,\gamma}
\end{aligned}$$

which completes the proof. ■

Proposition 6.18 *Let* $\beta_1 = \big(\mathbf{b}_1^{(1)}, \ldots, \mathbf{b}_n^{(1)}\big)$ *and* $\beta_2 = \big(\mathbf{b}_1^{(2)}, \ldots, \mathbf{b}_n^{(2)}\big)$ *be two bases for* $\mathbb{R}^n$*, and* $\gamma_1 = \big(\mathbf{c}_1^{(1)}, \ldots, \mathbf{c}_m^{(1)}\big)$ *and* $\gamma_2 = \big(\mathbf{c}_1^{(2)}, \ldots, \mathbf{c}_m^{(2)}\big)$ *be two bases for* $\mathbb{R}^m$*. Then, the change-of-coordinates rule for any linear transformation* $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ *is*

$$\boxed{[T]_{\beta_2,\gamma_2} = M_{\gamma_1,\gamma_2}[T]_{\beta_1,\gamma_1}M^{-1}_{\beta_1,\beta_2}}$$

In particular, if $m = n$ *and* $\beta_1 = \gamma_1 = \beta$ *and* $\beta_2 = \gamma_2 = \gamma$*,*

$$\boxed{[T]_\gamma = M_{\beta,\gamma}[T]_\beta M^{-1}_{\beta,\gamma}}$$

Proof: Again, this is merely an identity in terms of coordinate maps,

$$\begin{aligned}
M_{\gamma_1,\gamma_2}[T]_{\beta_1,\gamma_1}M^{-1}_{\beta_1,\beta_2} &= \Big[(\varphi_{\gamma_2} \circ \varphi_{\gamma_1}{}^{-1})\Big]_\sigma \Big[(\varphi_{\gamma_1} \circ T \circ \varphi_{\beta_1}{}^{-1})\Big]_\sigma \Big[(\varphi_{\beta_1} \circ \varphi^{-1}_{\beta_2})\Big]_\sigma \\
&= \Big[(\varphi_{\gamma_2} \circ \cancel{\varphi_{\gamma_1}}^{-1}) \circ (\cancel{\varphi_{\gamma_1}} \circ T \circ \cancel{\varphi_{\beta_1}}^{-1}) \circ (\cancel{\varphi_{\beta_1}} \circ \varphi^{-1}_{\beta_2})\Big]_\sigma \\
&= \Big[\varphi_{\gamma_2} \circ T \circ \varphi^{-1}_{\beta_2}\Big]_\sigma \\
&= [T]_{\beta_2,\gamma_2}
\end{aligned}$$

which completes the proof. ■

Proposition 6.19 *Consider three bases for* $\mathbb{R}^n$*,*

$$\beta = \big(\mathbf{b}_1, \ldots, \mathbf{b}_n\big), \quad \gamma = \big(\mathbf{c}_1, \ldots, \mathbf{c}_n\big), \quad \delta = \big(\mathbf{d}_1, \ldots, \mathbf{d}_n\big)$$

The three corresponding change-of-coordinates matrices are related by the following transformation rule, which can be permuted, of course:

$$\boxed{M_{\beta,\delta} = M_{\gamma,\delta}M_{\beta,\gamma}}$$

Proof: This is nothing but the observation

$$\varphi_\delta \circ \varphi_\beta^{-1} = \varphi_\delta \circ (\varphi_\gamma^{-1} \circ \varphi_\gamma) \circ \varphi_\beta^{-1} = (\varphi_\delta \circ \varphi_\gamma^{-1}) \circ (\varphi_\gamma \circ \varphi_\beta^{-1})$$

applied to I_n, or equivalently to each of its columns $\mathbf{e}_j$. ■

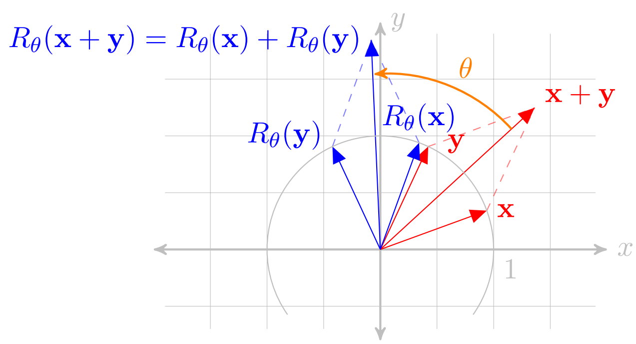
7 Examples of Coordinates in $\mathbb{R}^2$

7.1 Rotations in $\mathbb{R}^2$

Consider the following picture of the **unit circle** in $\mathbb{R}^2$,

$$S^1 = \{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\| = 1\} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \right\}$$

We rotate the two points $\mathbf{x}$ and $\mathbf{y}$ lying on it, as well as the sum $\mathbf{x}+\mathbf{y}$ lying outside of it, by an angle θ, that is we apply a rotation R_θ to them.



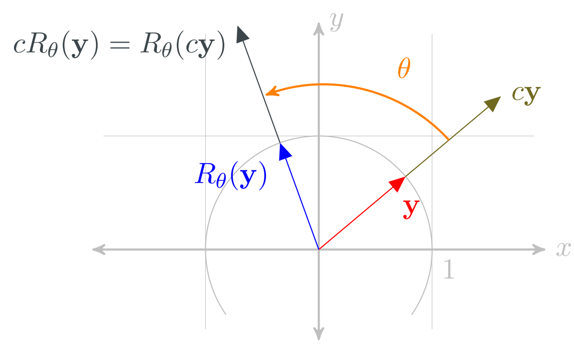
It definitely appears that

$$R_\theta(\mathbf{x}+\mathbf{y}) = R_\theta(\mathbf{x}) + R_\theta(\mathbf{y})$$

which is half of linearity—we could either rotate the point $\mathbf{x}+\mathbf{y}$ through an angle θ or we could rotate the vectors $\mathbf{x}$ and $\mathbf{y}$ on the unit circle S^1 through an angle θ and add the resulting vectors $R_\theta(\mathbf{x})$ and $R_\theta(\mathbf{y})$—and get the same result! The other half of linearity is preservation of scalar multiplication,

$$R_\theta(c\mathbf{x}) = cR_\theta(\mathbf{x})$$

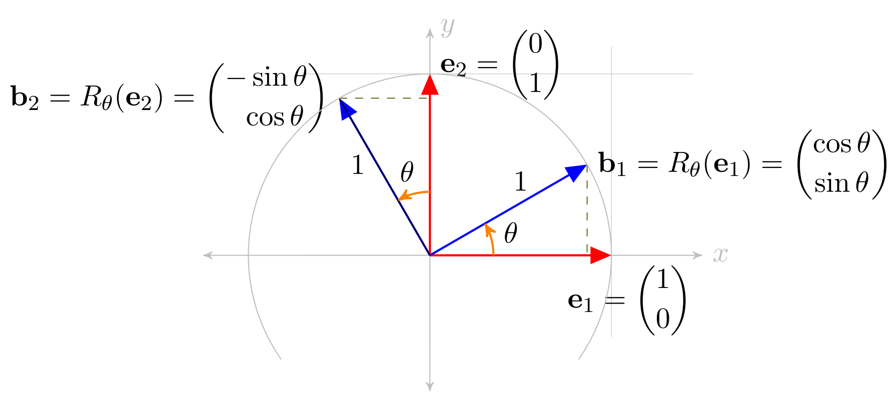
for all $c \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$:



Remark 7.1 Rather than be pedantic—for example, by defining rotations in a weird way, such as 'an isometry fixing one point' and then deducing linearity artificially from these supposedly more basic principles—let's take the linearity of rotations as given. If there is one thing trigonometry and plane geometry teach us, it is the linearity of rotations—linear isometries and projective functions are the modern way of encoding *congruence* and *similarity* of plane figures, *following Felix Klein's definitive recasting of classical geometry in terms of groups of shape-preservating operations, and its generalizations to other geometric properties* (second half of the 19th-c). Rotations are taken *a priori* to be linear, on philosophical grounds, *because* of the appeal of symbolic linear algebraic methods over the ancient constructive methods. After we introduce the generalized dot product, we'll return to these comments and consider the connections more attentively. ■

Remark 7.2 Rotations are one of *two types* of linear operators on $\mathbb{R}^2$ which *preserve lengths*. The other class is *reflections*. One easy way to distinguish between them is the fact that rotations leave only one vector unmoved (if $\theta \neq 2k\pi$, of course, otherwise $R_\theta = I_2$), namely the origin $\mathbf{0}$. Contrast this with reflections which leave an entire one-dimensional subspace unmoved, namely the line across which the reflection is occuring. More on this later. ■

Assuming the rotation R_θ to be linear, then, we can get a matrix representation of R_θ as follows: First, note that a rotation through a positive angle θ sends the standard basis vector $\mathbf{e}_1 = \langle 1, 0\rangle$ to $\langle \cos\theta, \sin\theta\rangle$, and it sends the other basis vector $\mathbf{e}_2 = \langle 0, 1\rangle$ to $\langle -\sin\theta, \cos\theta\rangle$:



Since

$$R_\theta(\mathbf{e}_1) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \qquad \text{and} \qquad R_\theta(\mathbf{e}_2) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

we have the standard matrix representation for R_θ,

$$[R_\theta]_\sigma \stackrel{\text{def}}{=} \Big([R_\theta(\mathbf{e}_1)]_\sigma \quad [R_\theta(\mathbf{e}_2)]_\sigma \Big) = \left(\left[\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}\right]_\sigma \quad \left[\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}\right]_\sigma \right) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Remark 7.3 We frequently just identify the linear operator $R_\theta \in \mathcal{L}(\mathbb{R}^n)$ with its standard matrix $[R_\theta]_\sigma$, and write

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

but it is important to remember the distinction. I will always be clear when I mean to contrast the operator with its standard representation. It is also clear that

$$\begin{aligned} R_\theta^{-1} &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}^{-1} = \frac{1}{\cos^2\theta + \sin^2\theta} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = R_{-\theta} \end{aligned}$$

since $\sin\theta$ is odd and $\cos\theta$ is even, $\sin(-\theta) = -\sin\theta$ and $\cos(-\theta) = \cos\theta$. ■

Once we have this, we can see what the effect is on any vector $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ in $\mathbb{R}^2$, by multiplying. Let us call the result $\mathbf{y} = \begin{pmatrix} u \\ v \end{pmatrix}$,

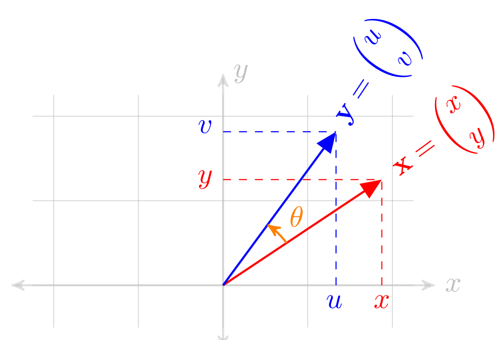
$$\mathbf{y} = R_\theta \mathbf{x}$$

and see what it looks like by carrying out this product:

$$\begin{pmatrix} u \\ v \end{pmatrix} = R_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}$$

Remark 7.4 This observation has ***two interpretations***:

(1) $\mathbf{x}$ is ***moved*** to a new position $\mathbf{y}$ in the plane. Both vectors are expressed in the same standard coordinate system.



(2) $\mathbf{x}$ is ***not moved***. Rather, $\mathbf{y} = \begin{pmatrix} u \\ v \end{pmatrix}$ is interpreted as **the same** $\mathbf{x}$, just **expressed in a different coordinate system** (namely γ, see below). Let

$$\beta = \big(\mathbf{b}_1, \mathbf{b}_2\big) = R_\theta \sigma = \big(R_\theta \mathbf{e}_1,\ R_\theta \mathbf{e}_2\big) = \left(\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}\right)$$

and

$$\gamma = \big(\mathbf{c}_1, \mathbf{c}_2\big) = R_{-\theta} \sigma = \big(R_{-\theta} \mathbf{e}_1,\ R_{-\theta} \mathbf{e}_2\big) = \left(\begin{pmatrix} \cos\theta \\ -\sin\theta \end{pmatrix}, \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix}\right)$$

be two bases, β rotating σ counterclockwise through θ radians, γ rotating σ clockwise. In terms of β, we interpret R_θ as the change-of-coordinates matrix

$$\boxed{R_\theta = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{pmatrix} = \begin{pmatrix} [\mathbf{b}_1]_\sigma & [\mathbf{b}_2]_\sigma \end{pmatrix} = M_{\beta,\sigma}}$$

and therefore $\boxed{R_{-\theta} = R_\theta^{-1} = M_{\sigma,\beta}}$. In terms of γ,

$$\boxed{R_{-\theta} = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{pmatrix} = \begin{pmatrix} [\mathbf{c}_1]_\sigma & [\mathbf{c}_2]_\sigma \end{pmatrix} = M_{\gamma,\sigma}}$$

and $\boxed{R_\theta = R_{-\theta}^{-1} = M_{\sigma,\gamma}}$. Using these, we *now we have to re-interpret* $\mathbf{x}$, for $\mathbf{x} \neq [\mathbf{x}]_\sigma$ here. Rather,

$$\boxed{\begin{aligned} \mathbf{x} &= R_\theta^{-1}\mathbf{y} = R_{-\theta}\mathbf{y} = M_{\sigma,\beta}[\mathbf{y}]_\sigma = [\mathbf{y}]_\beta \\ \mathbf{y} &= R_\theta \mathbf{x} = M_{\sigma,\gamma}[\mathbf{x}]_\sigma = [\mathbf{x}]_\gamma \end{aligned}}$$

Here is the picture:

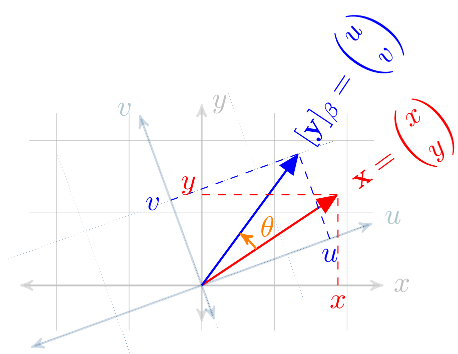


Figure (a): $\mathbf{x} = [\mathbf{y}]_\beta$

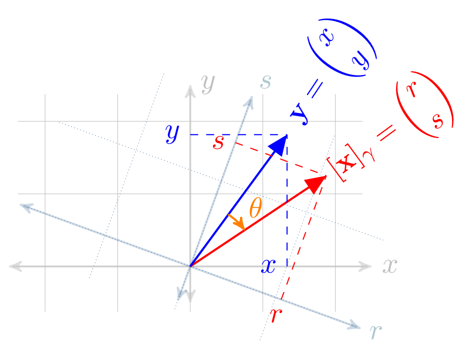


Figure (b): $\mathbf{y} = [\mathbf{x}]_\gamma$.

■

Exercise 7.5 Find the rotation matrices for the following angles: $\theta = 0$, $\theta = \pi/3$, $\theta = \pi/4$ and $\theta = 7\pi/12$. Use this to show that $R_{\pi/3}R_{\pi/4} = R_{7\pi/12}$. ■

The next exercise gives you the underlying reason for the result of the last problem.

Exercise 7.6 Use the angel sum formulas for $\sin(\theta+\varphi)$ and $\cos(\theta+\varphi)$ to demonstrate that, in general,

$$R_\theta R_\varphi = R_{\theta+\varphi}$$

for all $\theta, \varphi \in \mathbb{R}$. ■

Remark 7.7 Combined with our observation above that $R_\theta \in \mathrm{GL}(2,\mathbb{R})$, since

$$R_\theta^{-1} = R_{-\theta}$$

and your work in Exercise 7.5,

$$R_0 = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

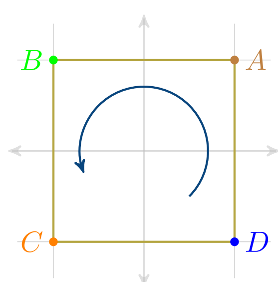
we come to see that the **set of all rotations** of the plane forms a group, called the **special orthogonal group**, denoted

$$\mathrm{SO}(2) = \left\{ R_\theta \mid \theta \in \mathbb{R} \right\}$$

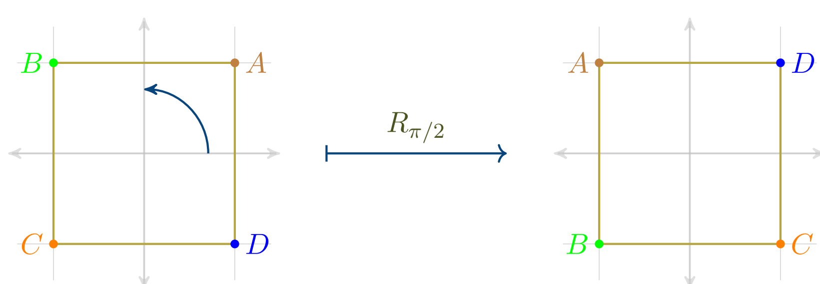
It is a proper **subgroup** of the set of *all* invertible 2×2 matrices $\mathrm{GL}(2,\mathbb{R})$, in the sense of being a sub*set* which is also a group (very similary to the way a subspace V or $\mathbb{R}^n$ is a subset which also a vector space). ■

Let us consider some subgroups of SO(2).

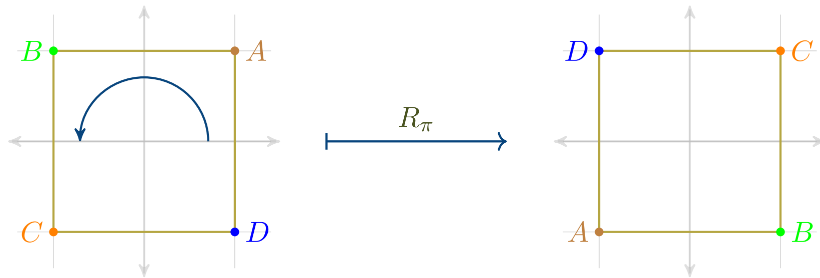
Example 7.8 Consider the subgroup of rotations which leave the unit square fixed, that is, which land the square back on itself:



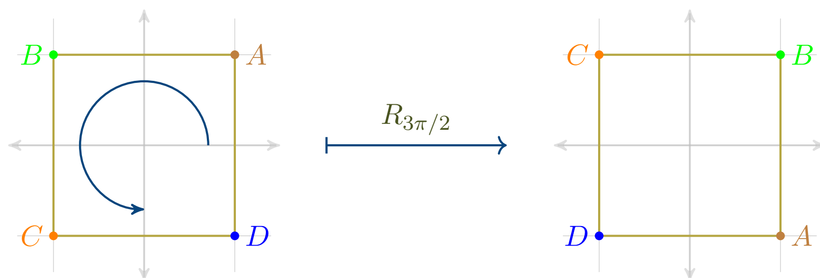
One can move vertices $A \to B \to C \to D$, which is a rotation through $\pi/2$



or one can move $A \to C$ and $B \to D$,



Finally, one can move $A \to D$, $B \to A$, $C \to B$ and $D \to C$:

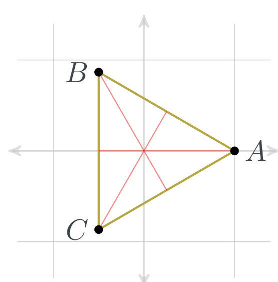


These three rotations, plus the identity 'rotation' I_2 through 0 (or more generally $2n\pi$) rdians, form the **rotation 4-group**,

$$\boxed{R_4 = \{I_2,\ R_{\pi/2},\ R_\pi,\ R_{3\pi/2}\}}$$

which is a subgroup of SO(2), since the inverse of any matrix in there is another matrix in there, and the product of any two matrices in there is another matrix in R_4. ■

Exercise 7.9 Find the rotation group R_3 which fixes the equilateral triangle centered at the origin.



Do the same with the pentagon, which should give you R_5. ■

Remark 7.10 This idea, of *invariance* of an object (a triangle, square, pentagon) under a *symmetry*, which is code for *group action*, is fundamental to modern math. It *redefines* symmetry as invariance under a group's action: the regularity part of the term **regular n-gon** refers, these days, to this property. Having found an equivalence between the *intuitive* idea of symmetry and the *conceptual formulation* in terms of an *invariance*, the modern mathematician switches from intuition to formalism immediately—which is not to say intuition thereby *disappears*! Rather, intuition is *codified* in a *formal language* for purposes of conceptual tidiness and computational implementability. Intuition lives on in the mind of the human, who *guides* the *creation* of new formlisms. If you have this goal in mind as you learn new math, you will understand these unstated assumptions and motivations, which lurk beneath the surface of any formal text. ■

Exercise 7.11 Let $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \in \mathbb{R}^2$ and consider the rotation $R_{\pi/6}$ through $\pi/6$. Show directly that $\mathbf{x} \cdot \mathbf{x}$ and $(R_{\pi/6}\mathbf{x}) \cdot (R_{\pi/6}\mathbf{x})$ are equal (to 10). ■

Exercise 7.12 Show that, for *any* $\theta \in \mathbb{R}$ and *any* vector $\mathbf{x} \in \mathbb{R}^2$, we have

$$\boxed{(R_\theta \mathbf{x}) \cdot (R_\theta \mathbf{x}) = \mathbf{x} \cdot \mathbf{x}}$$

(i.e. rotations preserve the dot product). Conclude that rotations preserve lengths,

$$\boxed{\|R_\theta \mathbf{x}\| = \|\mathbf{x}\|}$$

for all $\mathbf{x} \in \mathbb{R}^2$. ■

7.2 Projections in $\mathbb{R}^2$

Let ℓ be a line in $\mathbb{R}^2$ passing through the origin, a one-dimensional subspace. We want to project any arbitrary vector $\mathbf{x}$ in $\mathbb{R}^2$ onto ℓ. Facts about the dot product and its relation to projections are needed here, so we review them now (see the earlier notes, 'Points and Vectors in $\mathbb{R}^n$,' pages 8-10).

The next problem asks you to verify that, for a fixed $\mathbf{x} \in \mathbb{R}^n$, dotting with $\mathbf{x}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}$. Any linear transformation has a standard representation, as well as many other coordinate representations. Let's investigate these:

Exercise 7.13 (**Dotting-with-x-Transformation**) Prove parts (4) and (5) of Proposition 1.28, 'Points and Vectors in $\mathbb{R}^n$,' that for a fixed $\mathbf{x} \in \mathbb{R}^n$, the dot-with-$\mathbf{x}$ function

$$\boxed{\begin{aligned} &T_{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R} \\ &T_{\mathbf{x}}(\mathbf{v}) := \mathbf{x} \cdot \mathbf{v} \end{aligned}}$$

is *linear*: for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and all $c \in \mathbb{R}$,

$$\boxed{\begin{aligned} &T_{\mathbf{x}}(\mathbf{u} + \mathbf{v}) = T_{\mathbf{x}}(\mathbf{u}) + T_{\mathbf{x}}(\mathbf{v}) \\ &T_{\mathbf{x}}(c\mathbf{v}) = cT_{\mathbf{x}}(\mathbf{v}) \end{aligned}}$$

Therefore, $T_{\mathbf{x}} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$. Find also its standard representation, $[T_{\mathbf{x}}]_{\sigma_n,\sigma_1}$. ■

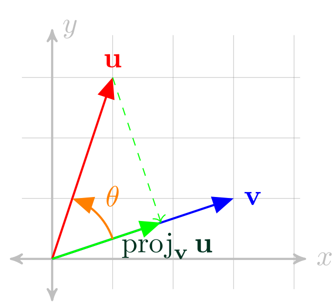
Exercise 7.14 Here is an illustration of the previous exercise with numbers: let $\mathbf{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \in \mathbb{R}^2$, and let $T_{\mathbf{x}} \in \mathcal{L}(\mathbb{R}^2, \mathbb{R})$ be the dotting-with-$\mathbf{x}$ transformation. Find its standard representation $[T_{\mathbf{x}}]_{\sigma_2,\sigma_1}$, as well as a nonstandard representation $[T_{\mathbf{x}}]_{\beta,\sigma_1}$, where

$$\beta = (\mathbf{b}_1, \mathbf{b}_2) = \left(\begin{pmatrix} -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 7 \\ 0 \end{pmatrix} \right)$$ ■

From Proposition 1.22 and the discussion on pages 16-17, 'Points and Vectors in $\mathbb{R}^n$,' we know that the **projection** of a vector $\mathbf{u}$ onto a vector $\mathbf{v}$ in $\mathbb{R}^n$ is given by the formula

$$\boxed{\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v}}$$

as in the following figure:



From this formula and the previous Exercise, 7.13, it is clear that, for fixed $\mathbf{v} \in \mathbb{R}^2$, the function

$$\mathrm{proj}_{\mathbf{v}} : \mathbb{R}^2 \to \mathbb{R}^2$$

is linear, so that $\mathrm{proj}_{\mathbf{v}} \in \mathcal{L}(\mathbb{R}^2)$.

Exercise 7.15 Let $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2$.

(1) Find the standard representation of $\mathrm{proj}_{\mathbf{v}} \in \mathcal{L}(\mathbb{R}^2)$,

$$[\mathrm{proj}_{\mathbf{v}}]_\sigma$$

(2) Now rotate $\mathbf{v}$ through $\pi/2$ to get an orthogonal vector, and look at the projection onto $\mathbf{w} = R_{\pi/2}\mathbf{v}$, $\mathrm{proj}_{R_{\pi/2}\mathbf{v}}$. This also, is a linear operator on $\mathbb{R}^2$. Find *its* standard representation,

$$[\mathrm{proj}_{R_{\pi/2}\mathbf{v}}]_\sigma$$

■

Remark 7.16 We can use these operators to decompose any vector $\mathbf{x} \in \mathbb{R}^2$ into **parallel** and **perpendicular components** to $\mathbf{v}$,

$$\boxed{\begin{aligned} \mathbf{x} &= \mathbf{x}_{\|\mathbf{v}} + \mathbf{x}_{\perp\mathbf{v}} \\ &= \mathrm{proj}_{\mathbf{v}}\,\mathbf{x} + \mathrm{proj}_{R_{\pi/2}\mathbf{v}}\,\mathbf{x} \\ &= [\mathrm{proj}_{\mathbf{v}}]_\sigma\mathbf{x} + [\mathrm{proj}_{R_{\pi/2}\mathbf{v}}]_\sigma\mathbf{x} \end{aligned}}$$

which is a useful idea, especially for the next section, when we look at reflections. If we reflect a vector $\mathbf{x}$ across a line ℓ through the origin, we only really flip the *perpendicular* component of $\mathbf{x}$ to ℓ: if $\ell = \mathrm{span}(\mathbf{v})$, it's simple,

$$R_\ell(\mathbf{x}) = R_\ell\big(\mathbf{x}_{\|\mathbf{v}} + \mathbf{x}_{\perp\mathbf{v}}\big) = \mathbf{x}_{\|\mathbf{v}} - \mathbf{x}_{\perp\mathbf{v}}$$

■

Exercise 7.17 Let's illustrate the idea with some numbers. Let $\mathbf{v} = \begin{pmatrix} -1 \\ 3 \end{pmatrix} \in \mathbb{R}^2$ be the basis for the line $\ell = \mathrm{span}(\mathbf{v})$, and let $\mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, as in Example 7.19 below. Decompose $\mathbf{x}$ into parallel and perpendicular components to $\mathbf{v}$,

$$\mathbf{x} = \mathbf{x}_{\|\mathbf{v}} + \mathbf{x}_{\perp\mathbf{v}}$$

and verify that the resulting reflection across ℓ is $\mathbf{y} = \mathbf{x}_{\|\mathbf{v}} - \mathbf{x}_{\perp\mathbf{v}} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$. ■

7.3 Reflections in $\mathbb{R}^2$

Definition 7.18 Let ℓ be a line in $\mathbb{R}^2$ passing through the origin, and let $\ell^\perp = R_{\pi/2}\ell$ be the line perpendicular to ℓ, which is easily achieved by rotating ℓ through $\pi/2$.[4] Each line is a subspace (of dimension 1), and so has a basis,

$$\beta = \big(\mathbf{b}_1\big) \qquad \text{(basis for } \ell\text{)}$$
$$\gamma = R_{\pi/2}\beta = \big(R_{\pi/2}(\mathbf{b}_1)\big) = \big(\mathbf{c}_1\big) \qquad \text{(basis for } \ell^\perp\text{)}$$

and consider a **reflection of the plane** across ℓ,

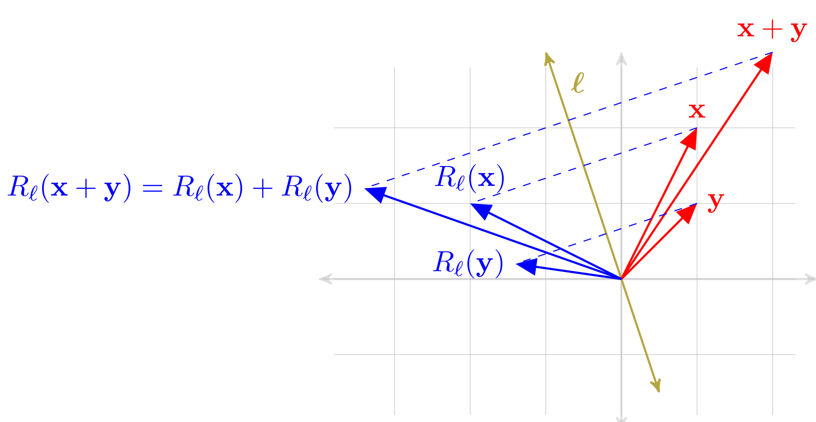
$$R_\ell(\mathbf{x}) = \begin{cases} \mathbf{x}, & \text{if } \mathbf{x} \in \ell \\ -\mathbf{x}, & \text{if } \mathbf{x} \in \ell^\perp \end{cases}$$

As with rotations, we take it as given that reflections are linear, for that is exactly what linear algebra was designed to capture about Euclidean geometry. We may take this as an axiom: ***Any reflection R_ℓ in $\mathbb{R}^2$ is a linear operator, a member of $\mathcal{L}(\mathbb{R}^2)$. In fact, R_ℓ is invertible, a member of $\mathrm{GL}(\mathbb{R}^2)$, for $R_\ell^2 = I_2$.***

$$R_\ell(\mathbf{x}+\mathbf{y}) = R_\ell(\mathbf{x}) + R_\ell(\mathbf{y}),$$
$$R_\ell(c\mathbf{x}) = cR_\ell(\mathbf{x})$$

■

This is the geometric content of the following figure:



To see more clearly the *algebraic* aspect of linearity, we need to represent $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}+\mathbf{y}$ and R_ℓ in a natural basis: let $\mathbf{b}_1$ be the basis vector for the line ℓ, and let $\mathbf{b}_2 = R_{\pi/2}\mathbf{b}_1$ be the rotated vector through $\pi/2$, the basis vector for $\ell^\perp$. Then

$$\beta = (\mathbf{b}_1, \mathbf{b}_2)$$

is a basis for $\mathbb{R}^2$.

[4] i.e. rotating every vector in ℓ by $\pi/2$

We can represent $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{x}+\mathbf{y}$ in β: given

$$\begin{aligned}\mathbf{x} &= a\mathbf{b}_1 + b\mathbf{b}_2\\ \mathbf{y} &= c\mathbf{b}_1 + d\mathbf{b}_2\\ \mathbf{x}+\mathbf{y} &= (a+c)\mathbf{b}_1 + (b+d)\mathbf{b}_2\end{aligned}$$

we have

$$[\mathbf{x}]_\beta = \begin{pmatrix} a\\ b\end{pmatrix}, \qquad [\mathbf{x}]_\beta = \begin{pmatrix} c\\ d\end{pmatrix}, \qquad [\mathbf{x}+\mathbf{y}]_\beta = \begin{pmatrix} a+c\\ b+d\end{pmatrix}$$

Also, with respect to this basis R_ℓ itself has a particularly simple representation: by its definition and our choices of $\mathbf{b}_i$,

$$\begin{aligned}R_\ell(\mathbf{b}_1) &= \mathbf{b}_1 = 1\mathbf{b}_1 + 0\mathbf{b}_2\\ R_\ell(\mathbf{b}_2) &= -\mathbf{b}_2 = 0\mathbf{b}_1 + (-1)\mathbf{b}_2\end{aligned} \implies \begin{aligned}[R_\ell]_\beta &= \begin{pmatrix}[R_\ell(\mathbf{b}_1)]_\beta & [R_\ell(\mathbf{b}_2)]_\beta\end{pmatrix}\\ &= \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\end{aligned}$$

Now apply Theorem 6.16, which says $[R_\ell(\mathbf{v})]_\beta = [R_\ell]_\beta[\mathbf{v}]_\beta$:

$$[R_\ell(\mathbf{v})]_\beta = [R_\ell]_\beta[\mathbf{v}]_\beta = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix} = \begin{pmatrix}a\\ -b\end{pmatrix}$$

which is, of course, exactly the statement

$$R_\ell(\mathbf{x}) = R_\ell(a\mathbf{b}_1 + b\mathbf{b}_2) = aR_\ell(\mathbf{b}_1) + bR_\ell(\mathbf{b}_2) = a\mathbf{b}_1 - b\mathbf{b}_2$$

Similarly,

$$\begin{aligned}R_\ell(\mathbf{y}) &= c\mathbf{b}_1 - d\mathbf{b}_2\\ R_\ell(\mathbf{x}+\mathbf{y}) &= (a+c)\mathbf{b}_1 - (b+d)\mathbf{b}_2\end{aligned}$$

and, <u>**if we *demand* that R_ℓ be *extended by linearity on* β**</u>, then additivity follows directly from vector addition and scalar multiplication rules, and is nothing but a special case of the proof of Theorem 4.1 above:

$$\begin{aligned}R_\ell(\mathbf{x}+\mathbf{y}) &= (a+b)\mathbf{b}_1 - (b+d)\mathbf{b}_2\\ &= \Big(a\mathbf{b}_1 - b\mathbf{b}_2\Big) + \Big(c\mathbf{b}_1 - d\mathbf{b}_2\Big)\\ &= R_\ell(\mathbf{x}) + R_\ell(\mathbf{y})\end{aligned}$$

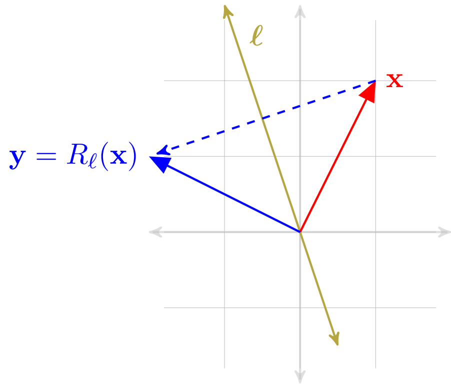
By the same logic, if c is any real number,

$$\begin{aligned}R_\ell(c\mathbf{x}) &= R_\ell\Big(c\big(a\mathbf{b}_1 + b\mathbf{b}_2\big)\Big)\\ &= R_\ell\big((ca)\mathbf{b}_1 + (cb)\mathbf{b}_2\big)\\ &= (ca)\mathbf{b}_1 - (cb)\mathbf{b}_2\\ &= c(a\mathbf{b}_1) - c(b\mathbf{b}_2)\\ &= c\big(a\mathbf{b}_1 - b\mathbf{b}_2\big)\\ &= cR_\ell(\mathbf{x})\end{aligned}$$

Example 7.19 Consider a reflection across the line

$$\ell = \left\{ \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; y = -3x \right\}$$

given in standard coordinates—remember, ℓ must be a subspace, so must contain $\mathbf{0}$, the origin.



Since the line is $y = -3x$, if we let $x = 1$, we get $y = -3$, and these are the coordinates of a basis vector

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

for ℓ. If we rotate this vector, say, by $\pi/2$, we get a basis vector for the orthogonal complement of ℓ, the line perpendicular:

$$\mathbf{b}_2 = \begin{pmatrix} -3 \\ -1 \end{pmatrix}$$

These two vectors form a basis $\beta = \big(\mathbf{b}_1, \mathbf{b}_2\big)$ for $\mathbb{R}^2$ in which the action of R_ℓ is transparent, from its definition:

$$\begin{aligned} R_\ell(\mathbf{b}_1) &= \ \ \mathbf{b}_1 \\ R_\ell(\mathbf{b}_2) &= -\mathbf{b}_2 \end{aligned}$$

so that R_ℓ has a simple β-representation, which follows from the above equations,

$$[R_\ell]_\beta = \Big([R_\ell(\mathbf{b}_1)]_\beta \ \ [R_\ell(\mathbf{b}_2)]_\beta \Big) = \Big([\mathbf{b}_1]_\beta \ \ [-\mathbf{b}_2]_\beta \Big) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Let us find its standard representation, using Proposition 6.18:

$$\begin{aligned} [R_\ell]_\sigma &= M_{\beta,\sigma}[R_\ell]_\beta M_{\beta,\sigma}^{-1} \\ &= \big(\mathbf{b}_1 \ \ \mathbf{b}_2\big)[R_\ell]_\beta\big(\mathbf{b}_1 \ \ \mathbf{b}_2\big)^{-1} \\ &= \begin{pmatrix} 1 & -3 \\ -3 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \left(-\frac{1}{10} \begin{pmatrix} -1 & 3 \\ 3 & 1 \end{pmatrix} \right) \\ &= -\frac{1}{5} \begin{pmatrix} 4 & 3 \\ 3 & -4 \end{pmatrix} \end{aligned}$$

You can verify this on the red vector, $\mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$:

$$[R_\ell(\mathbf{x})]_\sigma = [R_\ell]_\sigma [\mathbf{x}]_\sigma = -\frac{1}{5}\begin{pmatrix} 4 & 3 \\ 3 & -4 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = -\frac{1}{5}\begin{pmatrix} 10 \\ -5 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$ ■

Exercise 7.20 In the previous example, call the β-coordinates u and v, and write the equation $y = -3x$ in uv-coordinates. ■

Exercise 7.21 Find the β-representation of the resulting vector $\mathbf{y} = R_\ell(\mathbf{x})$ in the previous example. ■

Exercise 7.22 Find the standard representation of the reflection R_{ℓ_1} about the line ℓ_1 given by $y = 2x$ in standard coordinates. Then consider the reflection in Example 7.19 above, and call that R_{ℓ_2} (here, ℓ_2 is given by $y = -3x$ in σ-coordinates). Compute the standard representation of the composition $R_{\ell_2} \circ R_{\ell_1}$. It is a *rotation!* Find the angle of rotation. ■

Exercise 7.23 Show that the previous exercise holds true quite generally: **if R_{ℓ_1} and R_{ℓ_2} are *any* reflections of the plane, their composition $R_{\ell_2} \circ R_{\ell_1}$ is a rotation R_θ.** ■

Exercise 7.24 Let $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \in \mathbb{R}^2$ and consider the reflection R_ℓ across the line $\ell = \text{span}(\mathbf{v})$, where $\mathbf{v} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Show directly that $\mathbf{x} \cdot \mathbf{x}$ and $(R_\ell \mathbf{x}) \cdot (R_\ell \mathbf{x})$ are equal (to 10). ■

Exercise 7.25 Show that, for *any* line $\ell = \text{span}(\mathbf{v})$ in $\mathbb{R}^2$ and *any* vector $\mathbf{x} \in \mathbb{R}^2$, we have

$$\boxed{(R_\ell \mathbf{x}) \cdot (R_\ell \mathbf{x}) = \mathbf{x} \cdot \mathbf{x}}$$

(i.e. reflections preserve the dot product). Conclude that reflections preserve lengths,

$$\boxed{\|R_\ell \mathbf{x}\| = \|\mathbf{x}\|}$$

for all $\mathbf{x} \in \mathbb{R}^2$. ■

7.4 Dilations in $\mathbb{R}^2$

Fix a real number c, and consider the action of ***scaling*** vectors $\mathbf{x}$ in $\mathbb{R}^2$ by c,

$$T_c : \mathbb{R}^2 \to \mathbb{R}^2$$
$$T_c(\mathbf{x}) = c\mathbf{x}$$

By parts (1) and (3) of Theorem 1.11, T_c is a linear transformation, called a **dilation** of the plane:

$$T_c(\mathbf{x}+\mathbf{y}) = c(\mathbf{x}+\mathbf{y}) = c\mathbf{x} + c\mathbf{y} = T_c(\mathbf{x}) + T_c(\mathbf{y})$$
$$T_c(a\mathbf{x}) = c(a\mathbf{x}) = (ca)\mathbf{x} = (ac)\mathbf{x} = a(c\mathbf{x}) = aT_c(\mathbf{x})$$

What is the standard representation of T_c?

$$[T_c]_\sigma = \Big([T_c(\mathbf{e}_1)]_\sigma \;\; [T_c(\mathbf{e}_2)]_\sigma\Big) = \Big(c\mathbf{e}_1 \;\; c\mathbf{e}_2\Big) = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} = c\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = cI_2$$

which, upon reflection, is pretty obvious. Let us venture a representation in a different basis, say

$$\beta = \big(\mathbf{b}_1, \mathbf{b}_2\big) = \left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right)$$

Since $T_c(\mathbf{b}_i) = c\mathbf{b}_i$, we have

$$\begin{aligned}
[T_c]_\beta &= [T_c]_\beta \\
&= \Big([T_c(\mathbf{b}_1)]_\beta \;\; [T_c(\mathbf{b}_2)]_\beta\Big) \\
&= \Big([c\mathbf{b}_1]_\beta \;\; [c\mathbf{b}_2]_\beta\Big) \\
&= \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} \\
&= c\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= cI_2
\end{aligned}$$

which, again, is obvious. We could have deduced this from the change-of-coordinates formula, Proposition 6.18,

$$[T_c]_\beta = M_{\sigma,\beta}[T_c]_\sigma M_{\beta,\sigma}^{-1} = M_{\sigma,\beta}(cI_2)M_{\beta,\sigma}^{-1} = cM_{\sigma,\beta}M_{\beta,\sigma}^{-1} = cI_2 \tag{7.1}$$

It appears dilations look the same in all coordinate systems, since *the above equation holds for all bases* β, in fact. How do we explain this, and several other related questions?

7.4.1 Commuting Matrices: The Center of $\mathrm{GL}(n,\mathbb{R})$

The fact that $[T_c]_\beta = cI_2$ in all coordinate systems follows from equation (7.1) above, whose essence may be distilled into the simple statement

$$M_{\sigma,\beta}[T_c]_\sigma M_{\beta,\sigma}^{-1} = cI_2 = [T_c]_\sigma$$

Since any invertible matrix $B \in \mathrm{GL}(n,\mathbb{R})$ *is* a change-of-coordinates map $B = M_{\beta,\sigma}$, where $\beta = (\mathbf{b}_1,\dots,\mathbf{b}_n)$ is the basis consisting of its columns (pp. 20-21, 'Matrices and Linear Transformations'), the above identity may be rephrased

$$[T_c]_\sigma = B[T_c]_\sigma B^{-1}, \quad \text{for all } B \in \mathrm{GL}(n,\mathbb{R})$$

This idea then suggests another idea, equivalent to this one: commutativity of $[T_c]_\beta$ with all other $B \in \mathrm{GL}(n,\mathbb{R})$,

$$[T_c]_\sigma B = B[T_c]_\sigma, \quad \text{for all } B \in \mathrm{GL}(n,\mathbb{R})$$

which is clear anyway: $B(cI_n)B^{-1} = c(BI_nB^{-1}) = c(BB^{-1}) = cI_n$. OK, so **dilations commute with all invertible matrices**! The natural question arises, '*What are all the matrices* A *in* $\mathrm{GL}(n,\mathbb{R})$ *which commute with all other invertible matrices* B?',

$$AB = BA$$

Turns out *ony dilations* cI_n can do this! This is a theorem, stated as follows: ***dilations constitute the center of*** $\mathrm{GL}(n,\mathbb{R})$, which is algebra for

$$\begin{aligned} Z\big(\mathrm{GL}(n,\mathbb{R})\big) &\overset{\text{def}}{=} \{A \in \mathrm{GL}(n,\mathbb{R}) \mid AB = BA \text{ for all } B \in \mathrm{GL}(n,\mathbb{R})\} \\ &\overset{\text{a Theorem}}{=} \{cI_n \mid c \in \mathbb{R}\} \end{aligned}$$

The **center** of $\mathrm{GL}(n,\mathbb{R})$ (notice the Z, from 'zentrum'=center, *auf Deutch*) is by definition the set of all invertible matrices A satisfying $AB = BA$ for all $B \in \mathrm{GL}(n,\mathbb{R})$.

7.4.2 Dilations in Particular Directions: Eigenvalues and Eigenvectors

Given a square matrix $A \in M_n(\mathbb{R})$, are there *directions* in $\mathbb{R}^n$ in which A *acts as a scalar multiple of* I_n, $A\mathbf{x} = c\mathbf{x}$? We know that this *does actually happen*, for example

Example 7.26 The matrix $A = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix}$ acts as $-4I_2$,

$$\boxed{A = -4I_2}$$

but only in the direction of $\mathbf{x} = \begin{pmatrix} 6 \\ -5 \end{pmatrix}$:

$$\boxed{A\mathbf{x} = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix}\begin{pmatrix} 6 \\ -5 \end{pmatrix} = \begin{pmatrix} -24 \\ 20 \end{pmatrix} = -4\begin{pmatrix} 6 \\ -5 \end{pmatrix} = -4\mathbf{x}}$$

The scalar -4 is called an **eigenvalue** of A, and the direction vector $\mathbf{x}$ its associated **eigenvector** ('eigen-'=characteristic of it, or particular to it, *auf Deutch*). There is also a second direction eigenvector for this A, namely $\mathbf{y} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, but *its* eigenvector is different, it is 7:

$$\boxed{A\mathbf{y} = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 7 \end{pmatrix} = 7\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 7\mathbf{y}}$$

No other vectors except scalar multiples of* x *and* y *satisfy this property. E.g. $\mathbf{v} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$ is **not** a direction eigenvector:

$$A\mathbf{v} = \begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix}\begin{pmatrix} 3 \\ -2 \end{pmatrix} = \begin{pmatrix} -9 \\ 11 \end{pmatrix}$$

because $\begin{pmatrix} -9 \\ 11 \end{pmatrix}$ is not a scalar multiple of $\begin{pmatrix} 3 \\ -2 \end{pmatrix}$. ***However!*** These eigenvectors of A form a basis for $\mathbb{R}^2$:

$$\boxed{\beta = (\mathbf{x}, \mathbf{y}) = \left(\begin{pmatrix} 6 \\ -5 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)}$$

The β-representation of $\mathbf{v} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$ offers a solution:

$$[\mathbf{v}]_\beta = M_{\sigma,\beta}[\mathbf{v}]_\sigma = M_{\beta,\sigma}^{-1}\mathbf{v} = \begin{pmatrix} 6 & 1 \\ -5 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 3 \\ -2 \end{pmatrix} = \frac{1}{11}\begin{pmatrix} 1 & -1 \\ 5 & 6 \end{pmatrix}\begin{pmatrix} 3 \\ -2 \end{pmatrix} = \frac{1}{11}\begin{pmatrix} 5 \\ 3 \end{pmatrix}$$

for in this basis $\mathbf{v}$ decomposes into a ***sum of eigenvectors***:

$$\mathbf{v} = \frac{5}{11}\begin{pmatrix} 6 \\ -5 \end{pmatrix} + \frac{3}{11}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix} = \frac{5}{11}\mathbf{x} + \frac{3}{11}\mathbf{y}$$

and *therefore* the action of A on $\mathbf{v}$ becomes transparent:

$$\begin{aligned} A\mathbf{v} &= A\left(\frac{5}{11}\mathbf{x} + \frac{3}{11}\mathbf{y}\right) \\ &= \frac{5}{11}A\mathbf{x} + \frac{3}{11}A\mathbf{y} \\ &= \frac{5}{11}(-4\mathbf{x}) + \frac{3}{11}(7\mathbf{y}) \\ &= -\frac{20}{11}\mathbf{x} + \frac{21}{11}\mathbf{y} \end{aligned}$$

A couple of observations are in order:

- Imagine having to compute $A^{25}\mathbf{v}$ by first computing A^{25}: even for a 2×2 matrix like ours, it would be computationally taxing. Once you switch into β-coordinates, it's easy:

  $$A^{25}\mathbf{v} = A^{25}\left(\frac{5}{11}\mathbf{x} + \frac{3}{11}\mathbf{y}\right) = \frac{5}{11}A^{25}\mathbf{x} + \frac{3}{11}A^{25}\mathbf{y} = \frac{5}{11}(-4)^{25}\mathbf{x} + \frac{3}{11}7^{25}\mathbf{y}$$

  This calculation is easy. Now imagine *exponentiating* A,

  $$e^A \stackrel{\text{def}}{=} I_2 + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

  Applied to $\mathbf{v}$, we would have

  $$\boxed{\begin{aligned} e^A\mathbf{v} &= \sum_{n=0}^{\infty} \frac{1}{n!}A^n\mathbf{v} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!}A^n\left(\frac{5}{11}\mathbf{x} + \frac{3}{11}\mathbf{y}\right) \\ &= \frac{5}{11}\sum_{n=0}^{\infty} \frac{1}{n!}A^n\mathbf{x} + \frac{3}{11}\sum_{n=0}^{\infty} \frac{1}{n!}A^n\mathbf{y} \\ &= \frac{5}{11}\sum_{n=0}^{\infty} \frac{1}{n!}(-4)^n\mathbf{x} + \frac{3}{11}\sum_{n=0}^{\infty} \frac{1}{n!}7^n\mathbf{y} \\ &= \frac{5}{11}e^{-4}\mathbf{x} + \frac{3}{11}e^{7}\mathbf{y} \end{aligned}}$$

- Besides the ease of calculation, there is the more important philosophical idea of the ***representation*** of A—the principle involved here is Descartes' '*clarté*': From Exercise 3.8 we know that $\beta = (\mathbf{x}, \mathbf{y})$ is a basis for $\mathbb{R}^2$ iff

  $$\boxed{\mathbb{R}^2 = \operatorname{span}(\mathbf{x}) \oplus \operatorname{span}(\mathbf{y}) = \operatorname{span}\begin{pmatrix} 6 \\ -5 \end{pmatrix} \oplus \operatorname{span}\begin{pmatrix} 1 \\ 1 \end{pmatrix}}$$

  and in β-coordinates, A is very simple-looking: by Proposition 6.18

  $$\boxed{[A]_\beta \quad = \quad M_{\sigma,\beta}[A]_\sigma M_{\sigma,\beta}^{-1} \quad = \quad \frac{1}{11}\begin{pmatrix} 1 & -1 \\ 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 6 \\ 5 & 2 \end{pmatrix}\begin{pmatrix} 6 & 1 \\ -5 & 1 \end{pmatrix} \quad = \quad \begin{pmatrix} -4 & 0 \\ 0 & 7 \end{pmatrix}}$$

So $[A]_\beta = \begin{pmatrix} -4 & 0 \\ 0 & 7 \end{pmatrix}$ acts on $\mathbb{R}^2 = \text{span}(\mathbf{x}) \oplus \text{span}(\mathbf{y})$ by dilation on each component by the corresponding eigenvalue. We can usefully rephrase all this in terms of projections,

$$A = -4\,\text{proj}_{\mathbf{x}} + 7\,\text{proj}_{\mathbf{y}}$$

I don't like this notation. Let's switch to a slicker notation, $\pi_{\mathbf{x}}$ and $\pi_{\mathbf{y}}$:

$$\boxed{A = -4\pi_{\mathbf{x}} + 7\pi_{\mathbf{y}}}$$

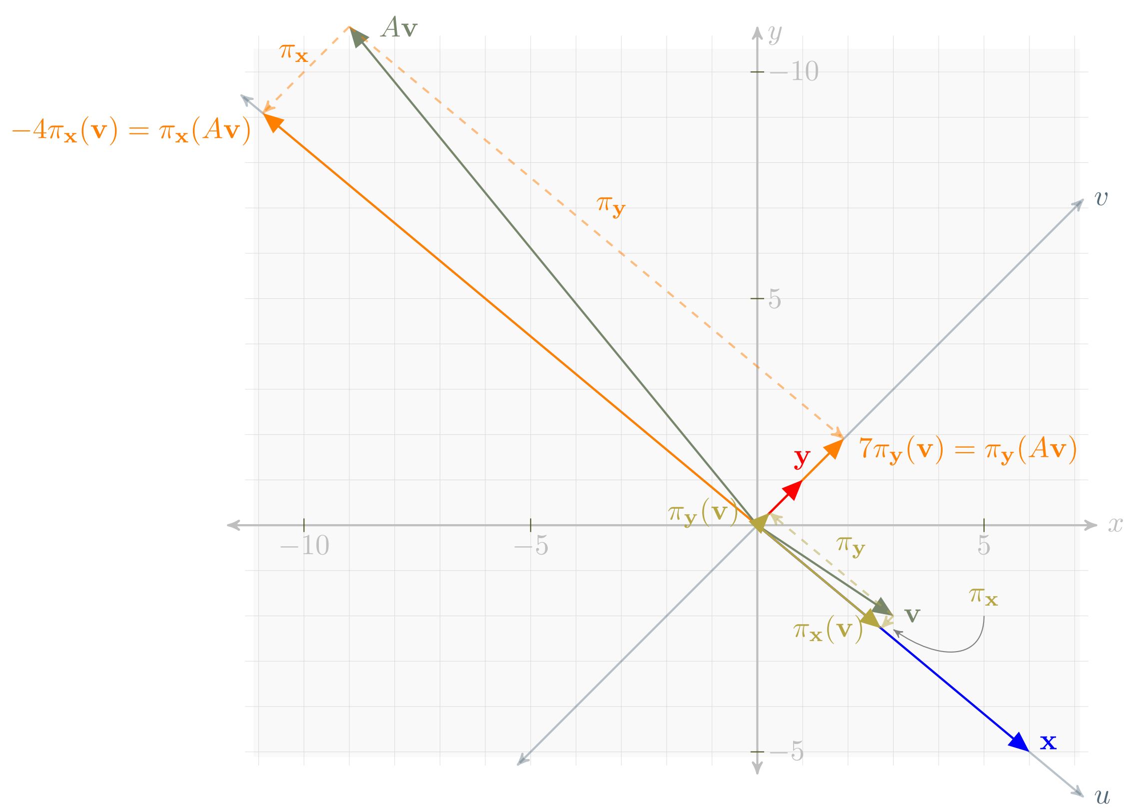
The above argument about $\mathbf{v}$ could then be neatly summed up like this:

$$\begin{aligned} A\mathbf{v} &= -4\pi_{\mathbf{x}}(\mathbf{v}) + 7\pi_{\mathbf{y}}(\mathbf{v}) \\ &= -4\left(\frac{5}{11}\mathbf{x}\right) + 7\left(\frac{3}{11}\mathbf{y}\right) \end{aligned}$$

Moreover, in $\mathbb{R}^2$ we have *polar coordinates*, so that any vector $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ can be written as a scalar multiple of the unit vector, which lies on the unit circle:

$$x^2$$

Here's the picture:



7.4.3 Dilations and Rotations in the Complex Plane

Let us now consider a completely different idea: the **complex plane** $\mathbb{C}$. The complex plane is just $\mathbb{R}^2$ as a set, but with a certain multiplication on the vectors, which we call complex numbers in this context. But there is a way to make the vector aspect explicit. The multiplication of two complex $z = a + bi$, $w = c + di$ follows form the fundamental identity $i^2 = -1$ on the imaginary number $i = \sqrt{-1}$, along with the ordinary rules of arithmetic:

$$\boxed{\begin{aligned} zw &= (a+bi)(c+di) \\ &= ac + adi + bci + bdi^2 \\ &= (ac - bd) + (ad + bc)i \end{aligned}}$$

The trick is to consider complex numbers as *matrix operators* on $\mathbb{R}^2$. This idea is called the **representation of** $\mathbb{C}$ **on** $\mathbb{R}^2$, and is gotten by means of a function (the *representation* itself):

$$\begin{aligned} &\rho : \mathbb{C} \to \mathcal{L}(\mathbb{R}^2) \\ &\rho(a+bi) := \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \end{aligned}$$

Observe how this representation respects the product:

$$\begin{aligned} \rho(z)\rho(w) &= \begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} c & -d \\ d & c \end{pmatrix} \\ &= \begin{pmatrix} ac - bd & -(ad+bc) \\ ad + bc & ac - bd \end{pmatrix} \\ &= \rho\big((ac-bd) + (ad+bc)i\big) \\ &= \rho(zw) \end{aligned}$$

The representation of the single imaginary number $i \in \mathbb{C}$ is the $\pi/2$-rotation matrix!

$$\rho(i) = \rho(0 + 1i) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = R_{\pi/2}$$

Of course,

$$R_{\pi/2}^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I_2$$

so, in a certain sense, ***rotations through $\pi/2$ are multiplication by i***! Now, *for a general complex number $z = a + ib$, the representation splits into a dilation plus a rotation through $\pi/2$ followed by a dilation*:

$$\rho(a+ib) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = aI + bR_{\pi/2}$$

But the rotation plus dilation is in fact all there is! Consider that $\mathbb{C}$ admits *polar coordinates*:

$$\boxed{a + bi = r(\cos\theta + i\sin\theta) = re^{i\theta}} \qquad \text{where } r = \sqrt{a^2 + b^2},$$

$$\theta = \begin{cases} \tan^{-1}(b/a), & \text{if } a \neq 0 \\ \pi/2, & \text{if } a = 0, b > 0 \\ -\pi/2, & \text{if } a = 0, b < 0 \end{cases}$$

θ is undefined at $a = b = 0$, but in that case we don't need θ becuse $r = 0$. Note that complex multiplication is easier to understand in this case:

$$zw = (re^{i\theta})(se^{i\varphi}) = rse^{i(\theta+\varphi)}$$

which is **a rotation of w through θ, followed by a dilation by r**. Indeed,

$$\rho(re^{i\theta}) = \rho(r\cos\theta + ir\sin\theta) = r\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} = rI_2 \circ R_\theta$$

Thus, the **unit length** complex numbers, those of the form $e^{i\theta}$ and lying on the circle $S^1 = \{z = a + bi \in \mathbb{C} \mid |z| = a^2 + b^2 = 1\}$, are in one-to-one correspondence with the *special linear group* SO(2): in this context, S^1 has special name, the **unitary group** $U(1)$:

$$\boxed{U(1) = \{e^{i\theta} \in \mathbb{C} \mid \theta \in \mathbb{R}\} \cong \mathrm{SO}(2)}$$

Indeed, exponentiation is present from the very beginning, since if we exponentiate the representation of $i\theta$ we get precisely the rotation matrix:

$$\begin{aligned}
e^{\rho(i\theta)} &= \exp\begin{pmatrix}0 & -\theta\\ \theta & 0\end{pmatrix}\\
&= \exp(\theta R_{\pi/2})\\
&= \sum_{n=0}^{\infty} \frac{1}{n!}(\theta R_{\pi/2})^n\\
&= \sum_{n=0}^{\infty} \frac{\theta^n}{n!} R_{\pi/2}^n\\
&= \sum_{k=0}^{\infty} \frac{(-1)^k\theta^{2k}}{(2k)!} I_2 + \sum_{k=0}^{\infty} \frac{(-1)^k\theta^{2k+1}}{(2k+1)!} R_{\pi/2}\\
&= \cos\theta I_2 + i\sin\theta R_{\pi/2}\\
&= \rho(e^{i\theta})\\
&= R_\theta
\end{aligned}$$

This is a curious fact. *We will consider the appearance of the exponential in 3-dimensional rotations as well!*