Matrix Representations of Linear Transformations and Changes of Coordinates

0.1 Subspaces and Bases

0.1.1 Definitions

A **subspace** V of $\mathbb{R}^n$ is a subset of $\mathbb{R}^n$ that contains the zero element and is closed under addition and scalar multiplication:

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

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

(3) $\mathbf{u} \in V$ and $k \in \mathbb{R} \implies k\mathbf{u} \in V$

Equivalently, V is a subspace if $a\mathbf{u} + b\mathbf{v} \in V$ for all $a, b \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v} \in V$. (You should try to prove that this is an equivalent statement to the first.)

Example 0.1 *Let $V = \{(t, 3t, -2t) \mid t \in \mathbb{R}\}$. Then V is a subspace of $\mathbb{R}^3$:*

(1) $\mathbf{0} \in V$ *because we can take $t = 0$.*

(2) *If $\mathbf{u}, \mathbf{v} \in V$, then $\mathbf{u} = (s, 3s, -2s)$ and $\mathbf{v} = (t, 3t, -2t)$ for some real numbers s and t. But then*

$$\mathbf{u} + \mathbf{v} = (s + t, 3s + 3t, -2s - 2t) = (s + t, 3(s + t), -2(s + t)) = (t', 3t', -2t') \in V$$

where $t' = s + t \in \mathbb{R}$.

(3) *If $\mathbf{u} \in V$, then $\mathbf{u} = (t, 3t, -2t)$ for some $t \in \mathbb{R}$, so if $k \in \mathbb{R}$, then*

$$k\mathbf{u} = (kt, 3(kt), -2(kt)) = (t', 3t', -2t') \in V$$

where $t' = kt \in \mathbb{R}$. ∎

Example 0.2 *The unit circle S^1 in $\mathbb{R}^2$ is not a subspace because it doesn't contain $\mathbf{0} = (0, 0)$ and because, for example, $(1, 0)$ and $(0, 1)$ lie in S but $(1, 0) + (0, 1) = (1, 1)$ does not. Similarly, $(1, 0)$ lies in S but $2(1, 0) = (2, 0)$ does not.* ∎

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

$$a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k \tag{0.1}$$

which is a vector in $\mathbb{R}^n$ (because $\mathbb{R}^n$ is a subspace of itself, right?). The $a_i \in \mathbb{R}$ are called the **coefficients** of the linear combination. If $a_1 = \cdots = a_k = 0$, then the linear combination is said to be **trivial**. In particular, considering the special case of $\mathbf{0} \in \mathbb{R}^n$, the zero vector, we note that $\mathbf{0}$ may always be represented as a linear combination of *any* vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$,

$$0\mathbf{u}_1 + \cdots + 0\mathbf{u}_k = \mathbf{0}$$

This representation is called the **trivial representation of $\mathbf{0}$** by $\mathbf{u}_1, \ldots, \mathbf{u}_k$. If, on the other hand, there are vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ and scalars $a_1, \ldots, a_n \in \mathbb{R}$ such that

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

where at least one $a_i \neq 0$, then that linear combination is called a **nontrivial representation of** $\mathbf{0}$. Using linear combinations we can generate subspaces, as follows. If S is a nonempty subset of $\mathbb{R}^n$, then the **span** of S is given by

$$\operatorname{span}(S) := \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{v} \text{ is a linear combination of vectors in } S\} \tag{0.2}$$

The span of the empty set, $\varnothing$, is by definition

$$\operatorname{span}(\varnothing) := \{0\} \tag{0.3}$$

Remark 0.3 *We showed in class that* $\operatorname{span}(S)$ *is always a subspace of* $\mathbb{R}^n$ *(well, we showed this for* S *a finite collection of vectors* $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$*, but you should check that it's true for any* S*).* ■

Let $V := \operatorname{span}(S)$ be the subspace of $\mathbb{R}^n$ spanned by some $S \subseteq \mathbb{R}^n$. Then S is said to **generate** or **span** V, and to be a **generating** or **spanning set** for V. If V is already known to be a subspace, then finding a spanning set S for V can be useful, because it is often easier to work with the smaller spanning set than with the entire subspace V, for example if we are trying to understand the behavior of linear transformations on V.

Example 0.4 *Let* S *be the unit circle in* $\mathbb{R}^3$ *which lies in the* x*-*y *plane. Then* $\operatorname{span}(S)$ *is the entire* x*-*y *plane.* ■

Example 0.5 *Let* $S = \{(x, y, z) \in \mathbb{R}^3 \mid x = y = 0,\ 1 < z < 3\}$*. Then* $\operatorname{span}(S)$ *is the* z*-axis.* ■

A nonempty subset S of a vector space $\mathbb{R}^n$ is said to be **linearly independent** if, taking any finite number of distinct vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in S$, we have for all $a_1, \ldots, a_k \in \mathbb{R}$ that

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

That is S is linearly independent if the *only* representation of $\mathbf{0} \in \mathbb{R}^n$ by vectors in S is the trivial one. In this case, the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ themselves are also said to be linearly independent. Otherwise, if there is at least one nontrivial representation of $\mathbf{0}$ by vectors in S, then S is said to be **linearly dependent**.

Example 0.6 *The vectors* $\mathbf{u} = (1, 2)$ *and* $\mathbf{v} = (0, -1)$ *in* $\mathbb{R}^2$ *are linearly independent, because if*

$$a\mathbf{u} + b\mathbf{v} = \mathbf{0}$$

that is

$$a(1, 2) + b(0, -1) = (0, 0)$$

then $(a, 2a - b) = (0, 0)$*, which gives a system of equations:*

$$\begin{aligned} a &= 0 \\ 2a - b &= 0 \end{aligned} \qquad \textit{or} \qquad \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

But the matrix $\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$ *is invertible, in fact it is its own inverse, so that left-multiplying both sides by it gives*

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}^2 \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which means $a = b = 0$*.* ■

Example 0.7 *The vectors* $(1,2,3),(4,5,6),(7,8,9) \in \mathbb{R}^3$ *are not linearly independent because*

$$1(1,2,3) - 2(4,5,6) + 1(7,8,9) = (0,0,0)$$

That is, we have found $a = 1$*,* $b = -2$ *and* $c = 1$*, not all of which are zero, such that* $a(1,2,3) + b(4,5,6) + c(7,8,9) = (0,0,0)$*.* ∎

Given $\varnothing \neq S \subseteq V$, a nonzero vector $\mathbf{v} \in S$ is said to be an **essentially unique linear combination** of the vectors in S if, up to order of terms, there is one and only one way to express $\mathbf{v}$ as a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_k \in S$. That is, if there are $a_1, \dots, a_n, b_1, \dots, b_\ell \in \mathbb{R}\backslash\{0\}$ and distinct $\mathbf{u}_1, \dots, \mathbf{u}_k \in S$ and distinct $\mathbf{v}_1, \dots, \mathbf{v}_\ell \in S$ distinct, then, re-indexing the b_is if necessary,

$$\left.\begin{aligned} \mathbf{v} &= a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_k \\ &= b_1\mathbf{v}_1 + \cdots + b_\ell\mathbf{v}_\ell \end{aligned}\right\} \quad\Longrightarrow\quad k = \ell \ \text{ and } \ \left\{\begin{aligned} a_i &= b_i \\ \mathbf{u}_i &= \mathbf{v}_i \end{aligned}\right\} \quad \text{for all } \ i = 1, \dots, k$$

If V is a subspace of $\mathbb{R}^n$, then a subset β of V is called a **basis** for V if it is linearly independent and spans V. We also say that the *vectors* of β form a basis for V. Equivalently, as explained in Theorem 0.11 below, β is a basis if every nonzero vector $\mathbf{v} \in V$ is an essentially unique linear combination of vectors in β.

Remark 0.8 *In the context of inner product spaces* V *of inifinite dimension, there is a difference between a vector space basis, the* **Hamel basis** *of* V*, and an orthonormal basis for* V*, the* **Hilbert basis** *for* V*, because though the two always exist, they are not always equal unless* $\dim(V) < \infty$*.* ∎

The **dimension** of a subspace V of $\mathbb{R}^n$ is the cardinality of any basis for V, i.e. the number of elements in β (which may in principle be infinite), and is denoted $\dim(V)$. This is a well defined concept, by Theorem 0.13 below, since all bases have the same size. V is **finite-dimensional** if it is the zero vector space $\{0\}$ or if it has a basis of finite cardinality. Otherwise, if it's basis has infinite cardinality, it is called **infinite-dimensional**. In the former case, $\dim(V) = |\beta| = k < \infty$ for some $n \in \mathbb{N}$, and V is said to be k**-dimensional**, while in the latter case, $\dim(V) = |\beta| = \kappa$, where κ is a cardinal number, and V is said to be κ**-dimensional**.

Remark 0.9 *Bases are not unique. For example,* $\beta = \{\mathbf{e}_1, \mathbf{e}_2\}$ *and* $\gamma = \{(1,1),(1,0)\}$ *are both bases for* $\mathbb{R}^2$*.* ∎

If V is finite-dimensional, say of dimension n, then an **ordered basis** for V a finite sequence or n-tuple $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ of linearly independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ such that $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for V. If V is infinite-dimensional but with a countable basis, then an ordered basis is a sequence $(\mathbf{v}_n)_{n\in\mathbb{N}}$ such that the set $\{\mathbf{v}_n \mid n \in \mathbb{N}\}$ is a basis for V.

0.1.2 Properties of Bases

Theorem 0.10 *Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ are linearly independent iff no $\mathbf{v}_i$ is a linear combination of the other $\mathbf{v}_j$.*

Proof: Let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ be linearly independent and suppose that $\mathbf{v}_k = c_1\mathbf{v}_1 + \cdots + c_{k-1}\mathbf{v}_{k-1}$ (we may suppose $\mathbf{v}_k$ is a linear combination of the other $\mathbf{v}_j$, else we can simply re-index so that this is the case). Then

$$c_1\mathbf{v}_1 + \cdots + c_{k-1}\mathbf{v}_{k-1} + (-1)\mathbf{v}_k = \mathbf{0}$$

But this contradicts linear independence, since $-1 \neq 0$. Hence $\mathbf{v}_k$ cannot be a linear combination of the other $\mathbf{v}_k$. By re-indexing the $\mathbf{v}_i$ we can conclude this for all $\mathbf{v}_i$.

Conversely, suppose $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly dependent, i.e. there are scalars $c_1, \ldots, c_k \in \mathbb{R}$ not all zero such that

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

Say $c_i \neq 0$. Then,

$$\mathbf{v}_i = \left(-\frac{c_1}{c_i}\right)\mathbf{v}_1 + \cdots + \left(-\frac{c_{i-1}}{c_i}\right)\mathbf{v}_{i-1} + \left(-\frac{c_{i+1}}{c_i}\right)\mathbf{v}_{i+1} + \cdots + \left(-\frac{c_k}{c_i}\right)\mathbf{v}_k$$

so that $\mathbf{v}_i$ is a linear combination of the other $\mathbf{v}_j$. This is the contrapositive of the equivalent statement, "If no $\mathbf{v}_i$ is a linear combination of the other $\mathbf{v}_j$, then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent." ∎

Theorem 0.11 *Let V be a subspace of $\mathbb{R}^n$. Then a collection $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for V iff every vector $\mathbf{v} \in V$ has an essentially unique expression as a linear combination of the basis vectors $\mathbf{v}_i$.*

Proof: Suppose β is a basis and suppose that $\mathbf{v}$ has two representations as a linear combination of the $\mathbf{v}_i$:

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \\ &= d_1\mathbf{v}_1 + \cdots + d_k\mathbf{v}_k \end{aligned}$$

Then,

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = (c_1 - d_1)\mathbf{v}_1 + \cdots + (c_k - d_k)\mathbf{v}_k$$

so by linear independence we must have $c_1 - d_1 = \cdots = c_k - d_k = 0$, or $c_i = d_i$ for all i, and so $\mathbf{v}$ has only one expression as a linear combination of basis vectors, up to order of the $\mathbf{v}_i$.

Conversely, suppose every $\mathbf{v} \in V$ has an essentially unique expression as a linear combination of the $\mathbf{v}_i$. Then clearly β is a spanning set for V, and moreover the $\mathbf{v}_i$ are linearly independent: for note, since $0\mathbf{v}_1 + \cdots + 0\mathbf{v}_k = \mathbf{0}$, by uniqueness of representations we must have $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0} \implies c_1 = \cdots = c_k = 0$. Thus β is a basis. ∎

Theorem 0.12 (Replacement Theorem) *Let V be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_p$ and $\mathbf{w}_1, \ldots, \mathbf{w}_q$ be vectors in V. If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are linearly independent and $\mathbf{w}_1, \ldots, \mathbf{w}_q$ span V, then $p \leq q$.*

Proof: Let $A = [\, \mathbf{w}_1 \ \cdots \ \mathbf{w}_q \,] \in \mathbb{R}^{n \times q}$ be the matrix whose columns are the $\mathbf{w}_j$ and let $B = [\, \mathbf{v}_1 \ \cdots \ \mathbf{v}_p \,] \in \mathbb{R}^{n \times p}$ be the matrix whose columns are the $\mathbf{v}_k$. Then note that

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \subseteq V = \operatorname{span}(\mathbf{w}_1, \ldots, \mathbf{w}_q) = \operatorname{im} A$$

Thus, there exist $\mathbf{u}_1, \ldots, \mathbf{u}_p \in \mathbb{R}^q$ such that $A\mathbf{u}_i = \mathbf{v}_i$. Consequently,

$$B = [\, \mathbf{v}_1 \ \cdots \ \mathbf{v}_p \,] = [\, A\mathbf{u}_1 \ \cdots \ \mathbf{u}_p \,] = A[\, \mathbf{u}_1 \ \cdots \ \mathbf{u}_p \,] = AC$$

where $C = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_p] \in \mathbb{R}^{q \times p}$. Now, since $\mathbf{v}_1 \ \cdots \ \mathbf{v}_p$ are linearly independent, $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$ implies all $c_i = 0$, i.e. $B\mathbf{c} = \mathbf{0}$ implies $\mathbf{c} = \mathbf{0}$, or $\ker B = \{\mathbf{0}\}$. But you will notice that $\ker C \subseteq \ker B$, since if $\mathbf{x} \in \ker C$, then the fact that $B = AC$ implies $B\mathbf{x} = (AC)\mathbf{x} = A(C\mathbf{x}) = A\mathbf{0} = \mathbf{0}$, or $\mathbf{x} \in \ker C$. Since $\ker B = \{\mathbf{0}\}$, this means that $\ker C = \{\mathbf{0}\}$ as well. But then C must have at least as many rows as columns, i.e. $p \leq q$, because $\operatorname{rref}(C)$ must have the form $\begin{bmatrix} I_p \\ O \end{bmatrix}$, possibly with no O submatrix, but at least with I_p in the top portion. ■

Theorem 0.13 *Let V be a subspace for $\mathbb{R}^n$. Then all bases for V have the same size.*

Proof: By the previous theorem two bases $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ and $\gamma = \{\mathbf{w}_1, \ldots, \mathbf{w}_q\}$ for V both span V and both are linearly independent, so we have $p \leq q$ and $p \geq q$. Therefore $p = q$. ■

Corollary 0.14 *All bases for $\mathbb{R}^n$ have n vectors.*

Proof: Notice that $\rho = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ forms a basis for $\mathbb{R}^n$: first, the elementary vectors $\mathbf{e}_i$ span $\mathbb{R}^n$, since if $\mathbf{x} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, then

$$\begin{aligned} \mathbf{x} = (a_1, \ldots, a_n) &= a_1(1, 0, \ldots, 0) + a_2(0, 1, 0, \ldots, 0) + \cdots + a_n(0, \ldots, 0, 1) \\ &= a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n \in \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_n) \end{aligned}$$

Also, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are linearly independent, for if

$$\mathbf{0} = c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = I_n\mathbf{c}$$

then $\mathbf{c} = (c_1, \ldots, c_n) \in \ker I_n = \{\mathbf{0}\}$, so $c_1 = \cdots = c_n = 0$. Since $|\rho| = n$, all bases β for $\mathbb{R}^n$ satisfy $|\beta| = n$ be the previous theorem. ■

Theorem 0.15 (Characterizations of Bases) *If V is a subspace of $\mathbb{R}^n$ and $\dim(V) = k$, then*

(1) *There are at most k linearly independent vectors in V. Consequently, a basis is a* **maximal linearly independent** *set in V.*

(2) *At least k vectors are needed to span V. Thus a basis is a* **minimal spanning set***.*

(3) *If k vectors in V are linearly independent, then they form a basis for V.*

(4) *If k vectors span V, then they form a basis for V.*

Proof: (1) If $\mathbf{v}_1, \ldots, \mathbf{v}_p \in V$ are linearly independent and $\mathbf{w}_1, \ldots, \mathbf{w}_k \in V$ form a basis for V, then $p \leq k$ by the Replacement Theorem. (2) If $\mathbf{v}_1, \ldots, \mathbf{v}_p \in V$ span V and $\mathbf{w}_1, \ldots, \mathbf{w}_k \in V$ form a basis for V then again we must have $k \leq p$ by the Replacement Theorem. (3) If $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$ are linearly independent, we must show they also span V. Pick $\mathbf{v} \in V$ and note that by (1) the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k, \sqsubseteq \in V$ are linearly dependent, because there are $k+1$ of them. (4) If $\mathbf{v}_1, \ldots, \mathbf{v}_k, \sqsubseteq \in V$ span V but are not linearly independent, then say $\mathbf{v}_k \in \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1})$. But in this case $V = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1})$, contradicting (2). ■

Theorem 0.16 *If* $A \in \mathbb{R}^{m \times n}$, *then* $\dim(\operatorname{im} A) = \operatorname{rank} A$.

Proof: This follows from Theorem 0.15 in *Systems of Linear Equations*, since if $B = \text{rref}(A)$, then $\operatorname{rank} A = \operatorname{rank} B = \#$ of columns of the form $\mathbf{e}_i$ in $B = \#$ of nonredundant vectors in A. ■

Theorem 0.17 (Rank-Nullity Theorem) *If* $A \in \mathbb{R}^{m \times n}$, *then*

$$\boxed{\dim(\ker A) + \dim(\operatorname{im} A) = n} \tag{0.4}$$

or

$$\boxed{\operatorname{null} A + \operatorname{rank} A = n} \tag{0.5}$$

Proof: If $B = \text{rref}(A)$, then $\dim(\ker A) = n - \#$ of leading 1s$= n - \operatorname{rank} A$. ■

0.2 Coordinate Representations of Vectors and Matrix Representations of Linear Transformations

0.2.1 Definitions

If $\beta = (\mathbf{v}_1, \ldots, \mathbf{v}_k)$ is an ordered basis for a subspace V of $\mathbb{R}^n$, then we know that for any vector $\mathbf{v} \in V$ there are unique scalars $a_1, \ldots, a_k \in \mathbb{R}$ such that

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

The **coordinate vector** of $\mathbf{v} \in \mathbb{R}^n$ with respect to, or relative to, β is defined to be the (column) vector in $\mathbb{R}^k$ consisting of the scalars a_i:

$$[\mathbf{x}]_\beta := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} \tag{0.6}$$

and the **coordinate map**, also called the **standard representation** of V with respect to β,

$$\phi_\beta : V \to \mathbb{R}^k \tag{0.7}$$

is given by

$$\phi_\beta(\mathbf{x}) = [\mathbf{x}]_\beta \tag{0.8}$$

Example 0.18 *Let $\mathbf{v} = (5, 7, 9) \in \mathbb{R}^3$ and let $\beta = (\mathbf{v}_1, \mathbf{v}_2)$ be the ordered basis for $V = \mathrm{span}(\mathbf{v}_1, \mathbf{v}_2)$, where $\mathbf{v}_1 = (1, 1, 1)$ and $\mathbf{v}_2 = (1, 2, 3)$. Can you express $\mathbf{v}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$? In other words, does $\mathbf{v}$ lie in V? If so, find $[\mathbf{v}]_\beta$.*

Solution: To find out whether $\mathbf{v}$ lies in V, we must see if there are scalars $a, b \in \mathbb{R}$ such that $\mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2$. Note, that if we treat $\mathbf{v}$ and the $\mathbf{v}_i$ as column vectors we get a matrix equation:

$$\mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2 = [\mathbf{v}_1 \;\; \mathbf{v}_2]\begin{bmatrix} a \\ b \end{bmatrix}$$

or

$$\begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix}$$

This is a system of equations, $A\mathbf{x} = \mathbf{b}$ with

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix}$$

Well, the agmented matrix $[A|\mathbf{b}]$ reduces to $\mathrm{rref}([A|\mathbf{b}])$ as follows:

$$\begin{bmatrix} 1 & 1 & 5 \\ 1 & 2 & 7 \\ 1 & 3 & 9 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

This means that $a = 3$ and $b = 2$, so that

$$\mathbf{v} = 3\mathbf{v}_1 + 2\mathbf{v}_2$$

and $\mathbf{v}$ lies in V, and moreover

$$\boxed{[\mathbf{v}]_\beta = \begin{bmatrix} 3 \\ 2 \end{bmatrix}} \qquad \blacksquare$$

In general, if $\beta = (\mathbf{v}_1, \dots, \mathbf{v}_k)$ is a basis for a subspace V of $\mathbb{R}^n$ and $\mathbf{v} \in V$, then the coordinate map will give us a matrix equation if we treat all the vectors $\mathbf{v}, \mathbf{v}_1, \dots, \mathbf{v}_k$ as column vectors:

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_k] \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}$$

or

$$\boxed{\mathbf{v} = B[\mathbf{v}]_\beta}$$

where $B = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_k]$, the $n \times k$ matrix with columns $\mathbf{v}_j$. If $V = \mathbb{R}^n$, then B will be an $n \times n$ matrix whose columns are linearly independent. Therefore, $\operatorname{im} B = \mathbb{R}^n$, so that by the Rank-Nullity Theorem $\ker B = \{\mathbf{0}\}$, which means B represents an injective and surjective linear transformation, and is therefore invertible. In this case, we can solve for $[\mathbf{v}]_\beta$ rather easily:

$$\boxed{[\mathbf{v}]_\beta = B^{-1}\mathbf{v}} \tag{0.9}$$

Let V and W be finite dimensional subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, with ordered bases $\beta = (\mathbf{v}_1, \dots, \mathbf{v}_k)$ and $\gamma = (\mathbf{w}_1, \dots, \mathbf{w}_\ell)$, respectively. If there exist (and there do exist) unique scalars $a_{ij} \in \mathbb{R}$ such that

$$T(\mathbf{v}_j) = \sum_{i=1}^{\ell} a_{ij}\mathbf{w}_i \qquad \text{for } \; j = 1, \dots, k \tag{0.10}$$

then the **matrix representation of a linear transformation** $T \in \mathcal{L}(V, W)$ **in the ordered bases** β **and** γ is the $\ell \times k$ matrix A defined by $A_{ij} = a_{ij}$,

$$\boxed{\begin{aligned} A = [T]_\beta^\gamma &:= \Big[\, [T(\mathbf{v}_1)]_\gamma \;\; [T(\mathbf{v}_2)]_\gamma \;\; \cdots \;\; [T(\mathbf{v}_k)]_\gamma \,\Big] \\ &= \left[\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{\ell 1} \end{pmatrix} \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{\ell 2} \end{pmatrix} \cdots \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{\ell k} \end{pmatrix} \right] = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{\ell 1} & \cdots & a_{\ell k} \end{pmatrix} \end{aligned}}$$

Note that $[T(\mathbf{v})]_\gamma = \varphi_\gamma(T(\mathbf{v})) = (\varphi_\gamma \circ T)(\mathbf{v})$.

Notation 0.19 *If $V = W$ and $\beta = \gamma$, we write $[T]_\beta$ instead of $[T]_\beta^\beta$.* ■

Example 0.20 *Let $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ be given by $T(x, y) = (x + y, 2x - y, 3x + 5y)$. In terms of matrices and column vectors T behaves as follows:*

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

But this matrix, call it A, is actually the representation of T with respect to the standard ordered bases $\rho_2 = (\mathbf{e}_1, \mathbf{e}_2)$ and $\rho_3 = (\mathbf{e}_1, \mathbf{e}_2, {}_3)$, that $A = [T]_{\rho_2}^{\rho_3}$. What if we were to choose different bases for $\mathbb{R}^2$ and $\mathbb{R}^3$? Say,

$$\beta = \big((1,1), (0,-1)\big), \qquad\qquad \gamma = \big((1,1,1), (1,0,1), (0,0,1)\big)$$

How would T look with respect to these bases? Let us first find the coordinate representations of $T(1,1)$ and $T(0,-1)$ with resepct to γ: Note, $T(1,1)=(2,1,8)$ and $T(0,-1)=(-1,1,-5)$, and to find $[(2,1,8)]_\gamma$ and $[(-1,1,-5)]$ we have to solve the equations:

$$\begin{bmatrix}1&1&0\\1&0&0\\1&1&1\end{bmatrix}\begin{bmatrix}2\\1\\8\end{bmatrix}_\gamma=\begin{bmatrix}2\\1\\8\end{bmatrix}\qquad \textit{and}\qquad \begin{bmatrix}1&1&0\\1&0&0\\1&1&1\end{bmatrix}\begin{bmatrix}-1\\1\\-5\end{bmatrix}_\gamma=\begin{bmatrix}-1\\1\\-5\end{bmatrix}$$

If B is the matrix $\begin{bmatrix}1&1&0\\1&0&0\\1&1&1\end{bmatrix}$, then $B^{-1}=\begin{bmatrix}0&1&0\\1&-1&0\\-1&0&1\end{bmatrix}$, so

$$\begin{bmatrix}2\\1\\8\end{bmatrix}_\gamma=\begin{bmatrix}0&1&0\\1&-1&0\\-1&0&1\end{bmatrix}\begin{bmatrix}2\\1\\8\end{bmatrix}=\begin{bmatrix}1\\1\\6\end{bmatrix}\qquad \textit{and}\qquad \begin{bmatrix}-1\\1\\-5\end{bmatrix}_\gamma=\begin{bmatrix}0&1&0\\1&-1&0\\-1&0&1\end{bmatrix}\begin{bmatrix}-1\\1\\-5\end{bmatrix}=\begin{bmatrix}1\\-2\\-4\end{bmatrix}$$

Let us verify this:

$$1(1,1,1)+1(1,0,1)+6(0,0,1)=(2,1,8)\qquad \textit{and}\qquad 1(1,1,1)-2(1,0,1)-4(0,0,1)=(-1,1,-5)$$

so indeed we have found $[T(1,1)]_\gamma$ and $[T(0,-1)]_\gamma$, and therefore

$$[T]_\beta^\gamma:=\Big[\,[T(1,1)]_\gamma\quad [T(0,-1)]_\gamma\,\Big]=\left[\begin{bmatrix}1\\1\\6\end{bmatrix}\quad\begin{bmatrix}1\\-2\\-4\end{bmatrix}\right]=\begin{bmatrix}1&1\\1&-2\\6&-4\end{bmatrix}\qquad\blacksquare$$

0.2.2 Properties of Coordinate and Matrix Representations

Theorem 0.21 (Linearity of Coordinates) *Let β be a basis for a subspace V of $\mathbb{R}^n$. Then, for all $\mathbf{x},\mathbf{y}\in V$ and all $k\in\mathbb{R}$ we have*

(1) $[\mathbf{x}+\mathbf{y}]_\beta=[\mathbf{x}]_\beta+[\mathbf{y}]_\beta$

(2) $[k\mathbf{x}]_\beta=k[\mathbf{x}]_\beta$

Proof: (1) On the one hand, $\mathbf{x}+\mathbf{y}=B[\mathbf{x}+\mathbf{y}]_\beta$, and on the other $\mathbf{x}=B[\mathbf{x}]_\beta$ and $\mathbf{y}=B[\mathbf{y}]_\beta$, so $\mathbf{x}+\mathbf{y}=B[\mathbf{x}]_\beta+B[\mathbf{y}]_\beta$. Thus,

$$B[\mathbf{x}+\mathbf{y}]_\beta=\mathbf{x}+\mathbf{y}=B[\mathbf{x}]_\beta+B[\mathbf{y}]_\beta=B([\mathbf{x}]_\beta+[\mathbf{y}]_\beta)$$

so that, subtracting the right hand side from both sides, we get

$$B\big([\mathbf{x}+\mathbf{y}]_\beta-([\mathbf{x}]_\beta+[\mathbf{y}]_\beta)\big)=\mathbf{0}$$

Now, B's columns are basis vectors, so they are linearly independent, which means $B\mathbf{x}=\mathbf{0}$ has only the trivial solution, because if $\beta=(\mathbf{v}_1,\dots,\mathbf{v}_k)$ and $\mathbf{x}=(x_1,\dots,x_k)$, then $\mathbf{0}=x_1\mathbf{v}_1+\cdots+x_k\mathbf{v}_k=B\mathbf{x}\implies x_1=\cdots=x_k=0$, or $\mathbf{x}=\mathbf{0}$. But this means the kernel of B is $\{\mathbf{0}\}$, so that

$$[\mathbf{x}+\mathbf{y}]_\beta-([\mathbf{x}]_\beta+[\mathbf{y}]_\beta)=\mathbf{0}$$

or

$$[\mathbf{x}+\mathbf{y}]_\beta=[\mathbf{x}]_\beta+[\mathbf{y}]_\beta$$

The proof of (2) follows even more straighforwardly: First, note that if $[\mathbf{x}]_\beta=[a_1\ \cdots\ a_k]^T$, then $\mathbf{x}=a_1\mathbf{v}_1+\cdots+a_k\mathbf{v}_k$, so that $k\mathbf{x}=ka_1\mathbf{v}_1+\cdots+ka_k\mathbf{v}_k$, and therefore

$$[k\mathbf{x}]_\beta=[ka_1\ \cdots\ ka_n]^T=k[a_1\ \cdots\ a_k]^T=k[\mathbf{x}]_\beta\qquad\blacksquare$$

Corollary 0.22 *The coordinate maps φ_β are linear, i.e. $\varphi_\beta \in \mathcal{L}(V,\mathbb{R}^k)$, and further they are isomorphisms, that is they are invertible, and so $\varphi_\beta \in \mathrm{GL}(V,\mathbb{R}^k)$.*

Proof: Linearity was shown in the previous theorem. To see that φ_β is an isomorphism, note first that φ_β takes bases to bases: if $\beta = (\mathbf{v}_1,\dots,\mathbf{v}_k)$ is a basis for V, then $\varphi_\beta(\mathbf{v}_i) = \mathbf{e}_i$, since $\mathbf{v}_i = 0\mathbf{v}_1 + \cdots + 1\mathbf{v}_i + \cdots + 0\mathbf{v}_k$. Thus, it takes β to the standard basis $\rho_k = (\mathbf{e}_1,\dots,\mathbf{e}_k)$ for $\mathbb{R}^k$. Consequently, it is surjective, because if $\mathbf{x} = (x_1,\dots,x_k) \in \mathbb{R}^k$, then

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_k\mathbf{e}_k = x_1\varphi_\beta(\mathbf{v}_1) + \cdots + x_k\varphi_\beta(\mathbf{v}_k) = \varphi_\beta(x_1\mathbf{v}_1 + \cdots + x_k\mathbf{v}_k)$$

If we let $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_k\mathbf{v}_k$, then we see that $\mathbf{v} \in V$ satisfies $\varphi_\beta(\mathbf{v}) = \mathbf{x}$. But φ_β is also injective: if $\mathbf{v} \in \ker\varphi_\beta$, then $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$, so that

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

By the linear independence of the $\mathbf{e}_i$ we must have $a_1 = \cdots = a_k = 0$, and so $\mathbf{v} = \mathbf{0}$. Thus, φ_β is also injective. ∎

Theorem 0.23 *Let V and W be finite-dimensional subspaces of $\mathbb{R}^n$ having ordered bases $\beta = (\mathbf{v}_1,\dots,\mathbf{v}_k)$ and $\gamma = (\mathbf{w}_1,\dots,\mathbf{w}_\ell)$, respectively, and let $T \in \mathcal{L}(V,W)$. Then for all $\mathbf{v} \in V$ we have*

$$[T(\mathbf{v})]_\gamma = [T]_\beta^\gamma [\mathbf{v}]_\beta \tag{0.11}$$

In other words, if $D = [T]_\beta^\gamma$ is the matrix representation of T in β and γ coordinates, with $T_D \in \mathcal{L}(\mathbb{R}^k,\mathbb{R}^\ell)$ the corresponding matrix multiplication map, if $A = [T]_{\rho_k}^{\rho_\ell}$ is the matrix representation of T in standard coordinates, with corresponding $T_A \in \mathcal{L}(\mathbb{R}^k,\mathbb{R}^\ell)$, and if $\phi_\beta \in \mathrm{GL}(V,\mathbb{R}^k)$ and $\phi_\gamma \in \mathrm{GL}(W,\mathbb{R}^\ell)$ are the respective coordinate maps, with matrix representations B^{-1} and C^{-1}, respectively, where $B = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_k \end{bmatrix}$ and $C = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_\ell \end{bmatrix}$, then

$$\phi_\gamma \circ T = T_D \circ \phi_\beta \qquad \textit{or, in terms of matrices,} \qquad C^{-1}A = DB^{-1} \tag{0.12}$$

and the following diagrams commute:

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle\phi_\beta}\downarrow & & \downarrow{\scriptstyle\phi_\gamma} \\ \mathbb{R}^k & \xrightarrow[T_D]{} & \mathbb{R}^\ell \end{array} \qquad \textit{or, in terms of matrices,} \qquad \begin{array}{ccc} V & \xrightarrow{A} & W \\ {\scriptstyle B^{-1}}\downarrow & & \downarrow{\scriptstyle C^{-1}} \\ \mathbb{R}^k & \xrightarrow[D]{} & \mathbb{R}^\ell \end{array}$$

Proof: If $\beta = (\mathbf{v}_1,\dots,\mathbf{v}_k)$ is an odered basis for V and $\gamma = (\mathbf{w}_1,\dots,\mathbf{w}_\ell)$ is an ordered basis for W, then let

$$[T]_\beta^\gamma = \Big[[T(\mathbf{v}_1)]_\gamma \ \cdots \ [T(\mathbf{v}_k)]_\gamma\Big] = \left[\begin{pmatrix} a_{11} \\ \vdots \\ a_{\ell 1}\end{pmatrix} \cdots \begin{pmatrix} a_{1kn} \\ \vdots \\ a_{\ell k}\end{pmatrix}\right] = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{\ell 1} & \cdots & a_{\ell k}\end{pmatrix}$$

Now, for all $\mathbf{u} \in V$ there are unique $b_1,\dots,b_n \in \mathbb{R}$ such that $\mathbf{u} = b_1\mathbf{v}_1 + \cdots + b_k\mathbf{v}_k$. Therefore,

$$\begin{aligned} T(\mathbf{u}) &= T(b_1\mathbf{v}_1 + \cdots + b_k\mathbf{v}_k) \\ &= b_1T(\mathbf{v}_1) + \cdots + b_kT(\mathbf{v}_k) \end{aligned}$$

so that by the linearity of φ_β,

$$\begin{aligned}
[T(\mathbf{u})]_\gamma = \phi_\gamma\big(T(\mathbf{u})\big) &= \phi_\gamma\big(b_1T(\mathbf{v}_1)+\cdots+b_kT(\mathbf{v}_k)\big)\\
&= b_1\phi_\gamma\big(T(\mathbf{v}_1)\big)+\cdots+b_k\phi_\gamma\big(T(\mathbf{v}_k)\big)\\
&= b_1[T(\mathbf{v}_1)]_\gamma+\cdots+b_k[T(\mathbf{v}_k)]_\gamma\\
&= \begin{bmatrix}[T(\mathbf{v}_1)]_\gamma & \cdots & [T(\mathbf{v}_k)]_\gamma\end{bmatrix}\begin{bmatrix}b_1\\ \vdots\\ b_k\end{bmatrix}\\
&= [T]_\beta^\gamma[\mathbf{u}]_\beta
\end{aligned}$$

This shows that $\phi_\gamma\circ T = T_D\circ\phi_\beta$, since $[T(\mathbf{u})]_\gamma = (\varphi_\gamma\circ T)(\mathbf{u})$ and $[T]_\beta^\gamma[\mathbf{u}]_\beta = (T_D\circ\varphi_\beta)(\mathbf{u})$. Finally, since $\varphi_\beta(\mathbf{x}) = B^{-1}\mathbf{x}$ and $\varphi_\gamma(\mathbf{y}) = C^{-1}\mathbf{y}$, we have the equivalent statement $C^{-1}A = DB^{-1}$. ■

Remark 0.24 *Two square matrices $A, B\in\mathbb{R}^{n\times n}$ are said to be* **similar***, denoted $A\sim B$, if there exists an invertible matrix $P\in\mathrm{GL}(n,\mathbb{R})$ such that*

$$A = PBP^{-1}$$

Similarity is an equivalence relation (it is reflexive, symmetric and transitive):

(1) (reflexivity) *$A\sim A$ because we can take $P = I_n$, so that $A = I_nAI_n^{-1}$*

(2) (symmetry) *If $A\sim B$, then there is an invertible P such that $A = PBP^{-1}$. But then left-multiplying by P^{-1} and right-multiplying P gives $P^{-1}AP = B$, so since P^{-1} is invertible, we have $B\sim A$.*

(3) (transitivity) *If $A\sim B$ and $B\sim C$, then there are invertible matrices P and Q such that $A = PBP^{-1}$ and $B = QCQ^{-1}$. Plugging the second into the first gives*

$$A = PBP^{-1} = P(QCQ^{-1})P^{-1} = (PQ)C(Q^{-1}P^{-1}) = (PQ)C(PQ)^{-1}$$

so since PQ is invertible, with inverse $Q^{-1}P^{-1}$, $A\sim C$.

In the previous theorem, if we take $W = V$ and $\gamma=\beta$, we'll have $B = C$, so that

$$B^{-1}A = DB^{-1}, \qquad \text{or} \qquad A = BDB^{-1}$$

which shows that $A\sim D$, i.e.

$$[T]_\rho\sim[T]_\beta$$

Since similarity is an equivalence relation, if β and γ are any two bases for V, then

$$[T]_\beta\sim[T]_\rho \quad \text{and} \quad [T]_\rho\sim[T]_\gamma \quad\Longrightarrow\quad [T]_\beta\sim[T]_\gamma$$

This demonstrates that if we represent a linear transformation $T\in\mathcal{L}(V,V)$ with respect to two different bases, then the corresponding matrices are similar

$$\boxed{[T]_\beta\sim[T]_\gamma}$$

Indeed, the invertible matrix P is $B^{-1}C$, because $[T]_\beta = B^{-1}[T]_\rho B$ and $[T]_\rho = C[T]_\gamma C^{-1}$, so that

$$[T]_\beta = B^{-1}[T]_\rho B = B^{-1}C[T]_\gamma C^{-1}B = (B^{-1}C)[T]_\gamma(B^{-1}C)^{-1}$$

We will show below that the converse is also true: if two matrices are similar, then they represent the same linear transformation, possibly with respect to different bases! ■

Theorem 0.25 *If V, W, Z are finite-dimensional subspaces of $\mathbb{R}^n$, $\mathbb{R}^m$ and $\mathbb{R}^p$, respectively, with ordered bases α, β and γ, respectively, and if $T \in \mathcal{L}(V,W)$ and $U \in \mathcal{L}(W,Z)$, then*

$$[U \circ T]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta \tag{0.13}$$

Proof: By the previous two theorems we have

$$\begin{aligned}
[U \circ T]_\alpha^\gamma &= \Big[[(U\circ T)(\mathbf{v}_1)]_\gamma \;\cdots\; [(U\circ T)(\mathbf{v}_k)]_\gamma \Big] \\
&= \Big[[U(T(\mathbf{v}_1))]_\gamma \;\cdots\; [U(T(\mathbf{v}_k))]_\gamma \Big] \\
&= \Big[[U]_\beta^\gamma [T(\mathbf{v}_1)]_\beta \;\cdots\; [U]_\beta^\gamma [T(\mathbf{v}_k)]_\beta \Big] \\
&= [U]_\beta^\gamma \Big[[T(\mathbf{v}_1)]_\beta \;\cdots\; [T(\mathbf{v}_k)]_\beta \Big] \\
&= [U]_\beta^\gamma [T]_\alpha^\beta
\end{aligned}$$

which completes the proof. ■

Corollary 0.26 *If V is an k-dimensional subspace of $\mathbb{R}^n$ with an ordered basis β and if $I \in \mathcal{L}(V,V)$ is the identity operator, then $[I]_\beta = I_k \in \mathbb{R}^{k\times k}$.*

Proof: Given any $T \in \mathcal{L}(V)$, we have $T = I \circ T$, so that $[T]_\beta = [I \circ T]_\beta = [I]_\beta [T]_\beta$, and similarly $[T]_\beta = [T]_\beta [I]_\beta$. Hence, taking $T = I$ we will have $[I]_\beta^2 = [I]_\beta$. Note also that $[I]_\beta^{-1} = [I]_\beta$, because $[I]_\beta^{-1}[I]_\beta = [I]_\beta [I]_\beta^{-1} = [I \circ I^{-1}]_\beta = [I]_\beta$. But then $[I]_\beta = I_n$, because any $A \in \mathbb{R}^{n\times n}$ that is invertible and satisfies $A^2 = A$ will satisfy $A = I_n$:

$$A = AI_n = A(AA^{-1}) = (AA)A^{-1} = A^2A^{-1} = AA^{-1} = I_k$$

An alternative proof follows directly from the definition, since $I(\mathbf{v}_i) = \mathbf{v}_i$, so that $[I(\mathbf{v}_i)]_\beta = \mathbf{e}_i$, whence $[I]_\beta = \big[\, [I(\mathbf{v}_1)]_\beta \;\cdots\; [I(\mathbf{v}_k)]_\beta \,\big] = [\mathbf{e}_1 \;\cdots\; \mathbf{e}_k] = I_k$. ■

Theorem 0.27 *If V and W are subspaces of $\mathbb{R}^n$ with ordered bases $\beta = (\mathbf{b}_1, \dots, \mathbf{b}_k)$ and $\gamma = (\mathbf{c}_1, \dots, \mathbf{c}_\ell)$, respectively, then the function*

$$\Phi : \mathcal{L}(V,W) \to \mathbb{R}^{\ell\times k} \tag{0.14}$$

$$\Phi(T) = [T]_\beta^\gamma \tag{0.15}$$

is an isomorphism, that is $\Phi \in \mathrm{GL}\big(\mathcal{L}(V,W),\ \mathbb{R}^{\ell\times k}\big)$, and consequently the space of linear transformations is isomorphic to the space of all $\ell \times k$ matrices:

$$\boxed{\mathcal{L}(V,W) \cong \mathbb{R}^{\ell\times k}} \tag{0.16}$$

Proof: First, Φ is linear: there exist scalars $r_{ij}, s_{ij} \in \mathbb{R}$, for $i = 1, \dots, \ell$ and $j = 1, \dots, k$, such that for any $T \in \mathcal{L}(V,W)$ we have

$$\begin{aligned}
T(\mathbf{b}_1) &= r_{11}\mathbf{c}_1 + \cdots + r_{\ell 1}\mathbf{c}_\ell & & & U(\mathbf{b}_1) &= s_{11}\mathbf{c}_1 + \cdots + s_{\ell 1}\mathbf{c}_\ell \\
&\vdots & &\text{and} & &\vdots \\
T(\mathbf{b}_k) &= r_{1k}\mathbf{c}_1 + \cdots + r_{\ell k}\mathbf{c}_\ell & & & U(\mathbf{b}_k) &= s_{1k}\mathbf{c}_1 + \cdots + s_{\ell k}\mathbf{c}_\ell
\end{aligned}$$

Hence, for all $s, t \in \mathbb{R}$ and $j = 1, \dots, n$ we have

$$(sT + tU)(b_j) = sT(b_j) + tU(b_j) = s\sum_{i=1}^{\ell} r_{ij}\mathbf{c}_i + t\sum_{i=1}^{\ell} s_{ij}\mathbf{c}_i = \sum_{i=1}^{\ell}(sr_{ij} + ts_{ij})\mathbf{c}_i$$

As a consequence of this and the rules of matrix addition and scalar multiplication we have

$$\begin{aligned}
\Phi(sT+tU) &= [sT+tU]_\beta^\gamma \\
&= \begin{pmatrix} sr_{11}+ts_{11} & \cdots & sr_{1k}+ts_{1k} \\ \vdots & \ddots & \vdots \\ sr_{\ell 1}+ts_{\ell 1} & \cdots & sr_{\ell k}+ts_{\ell k} \end{pmatrix} \\
&= s\begin{pmatrix} r_{11} & \cdots & r_{1k} \\ \vdots & \ddots & \vdots \\ r_{\ell 1} & \cdots & r_{\ell k} \end{pmatrix} + t\begin{pmatrix} s_{11} & \cdots & s_{1k} \\ \vdots & \ddots & \vdots \\ s_{\ell 1} & \cdots & s_{\ell k} \end{pmatrix} \\
&= s[T]_\beta^\gamma + t[U]_\beta^\gamma \\
&= s\Phi(T) + t\Phi(U)
\end{aligned}$$

Moreover, Φ is bijective, since for all $A \in \mathbb{R}^{\ell\times k}$ there is a unique linear transformation $T \in \mathcal{L}(V,W)$ such that $\Phi(T) = A$, because there exists a unique $T \in \mathcal{L}(V,W)$ such that

$$T(\mathbf{b}_j) = A_{1j}\mathbf{c}_1 + \cdots + A_{kj}\mathbf{c}_\ell \qquad \text{for } \; j = 1,\dots,k$$

This makes Φ onto, and also 1-1 because $\ker(\Phi) = \{T_0\}$, the zero operator, because for $O \in \mathcal{R}^{\ell\times k}$ there is only $T_0 \in \mathcal{L}(V,W)$ satisying $\Phi(T_0) = O$, because

$$T(\mathbf{b}_j) = 0\mathbf{c}_1 + \cdots + 0\mathbf{c}_\ell = 0 \qquad \text{for } \; j = 1,\dots,k$$

defines a unique transformation, T_0. ■

Theorem 0.28 *Let V and W be subspaces of $\mathbb{R}^n$ of the same dimension k, with ordered bases $\beta = (\mathbf{b}_1,\dots,\mathbf{b}_k)$ and $\gamma = (\mathbf{c}_1,\dots,\mathbf{c}_k)$, respectively, and let $T \in \mathcal{L}(V,W)$. Then T is an isomorphpism iff $[T]_\beta^\gamma$ is invertible, that is $T \in \mathrm{GL}(V,W)$, iff $[T]_\beta^\gamma \in \mathrm{GL}(k,\mathbb{R})$. In this case*

$$[T^{-1}]_\gamma^\beta = \big([T]_\beta^\gamma\big)^{-1} \tag{0.17}$$

Proof: If $T \in \mathrm{GL}(V,W)$ and $\dim(V) = \dim(W) = k$, then $T^{-1} \in \mathcal{L}(W,V)$, and $T \circ T^{-1} = I_W$ and $T^{-1} \circ T = I_V$, so that by Theorem 0.25 and Corollary 0.26 we have

$$[T]_\beta^\gamma[T^{-1}]_\gamma^\beta = [T\circ T^{-1}]_\gamma = [I_W]_\gamma = I_n = [I_V]_\beta = [T^{-1}\circ T]_\beta = [T^{-1}]_\gamma^\beta[T]_\beta^\gamma$$

so that $[T]_\beta^\gamma$ is invertibe with inverse $[T^{-1}]_\gamma^\beta$, and by the uniqueness of the multiplicative inverse in $\mathbb{R}^{k\times k}$, which follows from the uniqueness of $T_A^{-1} \in \mathcal{L}(\mathbb{R}^k,\mathbb{R}^k)$, we have

$$[T^{-1}]_\gamma^\beta = \big([T]_\beta^\gamma\big)^{-1}$$

Conversely, if $A = [T]_\beta^\gamma$ is invertible, there is a $n\times n$ matrix B such that $AB = BA = I_n$. Define $U \in \mathcal{L}(W,V)$ on the basis elements as follows, $U(\mathbf{c}_j) = \mathbf{v}_j = \sum_{i=1}^n B_{ij}\mathbf{b}_j$, and extend U by linearity. Then $B = [U]_\gamma^\beta$. To show that $U = T^{-1}$, note that

$$[U\circ T]_\beta = [U]_\gamma^\beta[T]_\beta^\gamma = BA = I_n = [I_V]_\beta \quad \text{and} \quad [T\circ U]_\gamma = [T]_\beta^\gamma[U]_\gamma^\beta = AB = I_n = [I_W]_\gamma$$

But since $\Phi \in \mathrm{GL}\big(\mathcal{L}(V,W),\mathbb{R}^{k\times k}\big)$ is an isomorphism, and therefore 1-1, we must have that

$$U\circ T = I_V \;\text{ and }\; T\circ U = I_W$$

By the uniqueness of the inverse, however, $U = T^{-1}$, and T is an isomorphism. ■

0.3 Change of Coordinates

0.3.1 Definitions

We now define the **change of coordinates**, or **change of basis, operator**. If V is a k-dimensional subspace of $\mathbb{R}^n$ and $\beta = (\mathbf{b}_1, \ldots, \mathbf{b}_k)$ and $\gamma = (\mathbf{c}_1, \ldots, \mathbf{c}_k)$ are two ordered bases for V, then the coordinate maps $\phi_\beta, \phi_\gamma \in \mathrm{GL}(V, \mathbb{R}^k)$, which are isomorphisms by Corollary 0.22 above, may be used to define a change of coordinates operator $\phi_{\beta,\gamma} \in \mathrm{GL}(\mathbb{R}^k, \mathbb{R}^k)$ changing β coordinates into γ coordinates, that is having the property

$$\phi_{\beta,\gamma}([\mathbf{v}]_\beta) = [\mathbf{v}]_\gamma \tag{0.18}$$

We define the operator as follows:

$$\phi_{\beta,\gamma} := \phi_\gamma \circ \phi_\beta^{-1} \tag{0.19}$$

The relationship between these three functions is illustrated in the following commutative diagram:

$$\begin{tikzcd}
\mathbb{R}^k \arrow[rr, "\phi_\gamma \circ \phi_\beta^{-1}"] & & \mathbb{R}^k \\
& V \arrow[ul, "\phi_\beta"] \arrow[ur, "\phi_\gamma"'] &
\end{tikzcd}$$

As we will see below, the change of coordinates operator has a matrix representation

$$M_{\beta,\gamma} = [\phi_{\beta,\gamma}]_\rho = [\phi_\gamma]_\beta^\rho = \begin{bmatrix} [\mathbf{b}_1]_\gamma & [\mathbf{b}_2]_\gamma & \cdots & [\mathbf{b}_n]_\gamma \end{bmatrix} \tag{0.20}$$

0.3.2 Properties of Change-of-Coordinate Maps and Matrices

Theorem 0.29 (Change of Coordinate Matrix) *Let V be a k-dimensional subspace of $\mathbb{R}^n$ and let $\beta = (\mathbf{b}_1, \ldots, \mathbf{b}_k)$ and $\gamma = (\mathbf{c}_1, \ldots, \mathbf{c}_k)$ be two ordered bases for V. Since ϕ_β and ϕ_γ are isomorphisms, the following diagram commutes,*

$$\begin{tikzcd}
\mathbb{R}^k \arrow[rr, "\phi_\gamma \circ \phi_\beta^{-1}"] & & \mathbb{R}^k \\
& V \arrow[ul, "\phi_\beta"] \arrow[ur, "\phi_\gamma"'] &
\end{tikzcd}$$

and the change of basis operator $\phi_{\beta,\gamma} := \phi_\gamma \circ \phi_\beta^{-1} \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^k)$, changing β coordinates into γ coordinates, is an isomorphism. It's matrix representation,

$$M_{\beta,\gamma} = [\phi_{\beta,\gamma}]_\rho \in \mathbb{R}^{k\times k} \tag{0.21}$$

where $\rho = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is the standard ordered basis for $\mathbb{R}^k$, is called the **change of coordinate matrix***, and it satisfies the following conditions:*

1. $$\begin{aligned} M_{\beta,\gamma} &= [\phi_{\beta,\gamma}]_\rho \\ &= [\phi_\gamma]_\beta^\rho \\ &= \begin{bmatrix} [\mathbf{b}_1]_\gamma & [\mathbf{b}_2]_\gamma & \cdots & [\mathbf{b}_k]_\gamma \end{bmatrix} \end{aligned}$$
2. $[\mathbf{v}]_\gamma = M_{\beta,\gamma}[\mathbf{v}]_\beta, \quad \forall \mathbf{v} \in V$
3. $M_{\beta,\gamma}$ *is invertible and* $\quad M_{\beta,\gamma}^{-1} = M_{\gamma,\beta} = [\phi_\beta]_\gamma^\rho$

Proof: The first point is shown as follows:

$$\begin{aligned}
M_{\beta,\gamma} = [\phi_{\beta,\gamma}]_\rho &= \Big[\phi_{\beta,\gamma}(\mathbf{e}_1)\ \ \phi_{\beta,\gamma}(\mathbf{e}_2)\ \cdots\ \phi_{\beta,\gamma}(\mathbf{e}_k)\Big] \\
&= \Big[(\phi_\gamma \circ \phi_\beta^{-1})(\mathbf{e}_1)\ \ (\phi_\gamma \circ \phi_\beta^{-1})(\mathbf{e}_2)\ \cdots\ (\phi_\gamma \circ \phi_\beta^{-1})(\mathbf{e}_k)\Big] \\
&= \Big[(\phi_\gamma \circ \phi_\beta^{-1})([\mathbf{b}_1]_\beta)\ \ (\phi_\gamma \circ \phi_\beta^{-1})([\mathbf{b}_2]_\beta)\ \cdots\ (\phi_\gamma \circ \phi_\beta^{-1})([\mathbf{b}_k]_\beta)\Big] \\
&= \Big[\phi_\gamma(\mathbf{b}_1)\ \ \phi_\gamma(\mathbf{b}_2)\ \cdots\ \phi_\gamma(\mathbf{b}_k)\Big] \\
&= \Big[[\mathbf{b}_1]_\gamma\ \ [\mathbf{b}_2]_\gamma\ \cdots\ [\mathbf{b}_n]_\gamma\Big] \\
&= [\phi_\gamma]_\beta
\end{aligned}$$

or, alternatively, by Theorem 0.25 and Theorem 0.28 we have that

$$M_{\beta,\gamma} = [\phi_{\beta,\gamma}]_\rho = [\phi_\gamma \circ \phi_\beta^{-1}]_\rho = [\phi_\gamma]_\beta^\rho[\phi_\beta^{-1}]_\rho^\beta = [\phi_\gamma]_\gamma^\rho([\phi_\beta]_\beta^\rho)^{-1} = [\phi_\gamma]_\beta^\rho I_n^{-1} = [\phi_\gamma]_\beta^\rho I_n = [\phi_\gamma]_\beta^\rho$$

The second point follows from Theorem 0.23, since

$$\phi_\gamma(\mathbf{v}) = (\phi_\gamma \circ I)(\mathbf{v}) = (\phi_\gamma \circ (\phi_\beta^{-1} \circ \phi_\beta))(\mathbf{v}) = ((\phi_\gamma \circ \phi_\beta^{-1}) \circ \phi_\beta)(\mathbf{v})$$

implies that

$$[\mathbf{v}]_\gamma = [\phi_\gamma(\mathbf{v})]_\rho = [((\phi_\gamma \circ \phi_\beta^{-1}) \circ \phi_\beta)(\mathbf{v})]_\rho = [\phi_\gamma \circ \phi_\beta^{-1}]_\rho[\phi_\beta(\mathbf{v})]_\rho = [\phi_{\beta,\gamma}]_\rho[\mathbf{v}]_\beta = M_{\beta,\gamma}[\mathbf{v}]_\beta$$

And the last point follows from the fact that ϕ_β and ϕ_γ are isomorphism, so that $\phi_{\beta,\gamma}$ is an isomorphism, and hence $\phi_{\beta,\gamma}^{-1} \in \mathcal{L}(\mathbb{R}^k)$ is an isomorphism, and because the diagram above commutes we must have

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

so that by (1)

$$M_{\beta,\gamma}^{-1} = [\phi_{\beta,\gamma}^{-1}]_\rho = [\phi_{\gamma,\beta}]_\rho = M_{\gamma,\beta}$$

or alternatively by Theorem 0.28

$$M_{\beta,\gamma}^{-1} = ([\phi_{\beta,\gamma}]_\rho)^{-1} = [\phi_{\beta,\gamma}^{-1}]_\rho = [\phi_{\gamma,\beta}]_\rho = [\phi_\beta]_\gamma^\rho = M_{\gamma,\beta}$$ ■

Corollary 0.30 (Change of Basis) *Let V and W be subspaces of $\mathbb{R}^n$ and let (β, γ) and (β', γ') be pairs of ordered bases for V and W, respectively. If $T \in \mathcal{L}(V, W)$, then*

$$[T]_{\beta'}^{\gamma'} = M_{\gamma,\gamma'}[T]_\beta^\gamma M_{\beta',\beta} \tag{0.22}$$

$$= M_{\gamma,\gamma'}[T]_\beta^\gamma M_{\beta,\beta'}^{-1} \tag{0.23}$$

where $M_{\gamma,\gamma'}$ and $M_{\beta',\beta}$ are change of coordinate matrices. That is, the following diagram commutes

$$\begin{tikzcd}
\mathbb{R}^k \arrow[rrr, "T_A"] \arrow[dr, "\phi_{\beta,\beta'}"] & & & \mathbb{R}^k \arrow[dl, "\phi_{\gamma',\gamma}"'] \\
& \mathbb{R}^k \arrow[r, "T_{A'}"] & \mathbb{R}^k & \\
& V \arrow[r, "T"'] \arrow[u, "\phi_{\beta'}"'] \arrow[uul, "\phi_\beta"] & W \arrow[u, "\phi_{\gamma'}"] \arrow[uur, "\phi_\gamma"'] &
\end{tikzcd}$$

Proof: This follows from the fact that if $\beta = (\mathbf{b}_1, \ldots, \mathbf{b}_k)$, $\beta' = (\mathbf{b}'_1, \ldots, \mathbf{b}'_k)$, $\gamma = (\mathbf{c}_1, \ldots, \mathbf{c}_\ell)$, $\gamma = (\mathbf{c}'_1, \ldots, \mathbf{c}'_\ell)$, then for each $i = 1, \ldots, k$ we have

$$[T(\mathbf{b}'_i)]_{\gamma'} = [(\phi_{\gamma,\gamma'} \circ T \circ \phi^{-1}_{\beta,\beta'})(\mathbf{b}'_i)]_{\gamma'}$$

so

$$\begin{aligned}
[T]^{\gamma'}_{\beta'} &= [\phi_{\gamma,\gamma'} \circ T \circ \phi^{-1}_{\beta,\beta'}]^{\gamma'}_{\beta'} \\
&= [\phi_{\gamma,\gamma'}]_{\rho_m} [T]^{\gamma}_{\beta} [\phi^{-1}_{\beta,\beta'}]_{\rho_n} \\
&= [\phi_{\gamma,\gamma'}]_{\rho_m} [T]^{\gamma}_{\beta} ([\phi_{\beta,\beta'}]_{\rho_n})^{-1} \\
&= M_{\gamma,\gamma'} [T]^{\gamma}_{\beta} M^{-1}_{\beta,\beta'}
\end{aligned}$$

which completes the proof. ■

Corollary 0.31 (Change of Basis for a Linear Operator) *If V is a subspaces of $\mathbb{R}^n$ with ordered bases β and γ, and $T \in \mathcal{L}(V)$, then*

$$[T]_\gamma = M_{\beta,\gamma} [T]_\beta M^{-1}_{\beta,\gamma} \tag{0.24}$$

where $M_{\beta,\gamma}$ is the change of coordinates matrix. ■

Corollary 0.32 *If we are given any two of the following:*

(1) *$A \in \mathbb{R}^n$ invertible*

(2) *an ordered basis β for $\mathbb{R}^n$*

(3) *an ordered basis γ for $\mathbb{R}^n$*

The third is uniquely determined by the equation $A = M_{\beta,\gamma}$, where $M_{\beta,\gamma}$ is the change of coordinates matrix of the previous theorem.

Proof: If we have $A = M_{\beta,\gamma} = [\phi_\beta]_\gamma = \Big[[\mathbf{b}_1]_\gamma \;\; [\mathbf{b}_2]_\gamma \;\; \cdots \;\; [\mathbf{b}_n]_\gamma\Big]$, suppose we know A and γ. Then, $[\mathbf{b}_i]_\gamma$ is given by A, so $\mathbf{b}_i = A_{i1}\mathbf{c}_1 + \cdots + A_{in}\mathbf{c}_n$, so β is uniquely determined. If β and γ are given, then by the previous theorem $M_{\beta,\gamma}$ is given by $M_{\beta,\gamma} = \Big[[\mathbf{b}_1]_\gamma \;\; [\mathbf{b}_2]_\gamma \;\; \cdots \;\; [\mathbf{b}_n]_\gamma\Big]$. Lastly, if A and β are given, then γ is given by the first case applied to $A^{-1} = M_{\gamma,\beta}$. ■