Lecture 7: Linear transformations

Let V and W be two vector spaces over a field K. We say that a transformation

$$F : V \mapsto W \tag{1}$$

is *linear* if

1. $F(u+v) = F(u) + F(v) \qquad \forall u, v \in V,$
2. $F(cu) = cF(u) \qquad \forall u \in V, \quad \forall c \in K.$

Conditions 1. and 2. imply that

$$F(au+bv) = aF(u) + bF(v) \qquad \forall u, v \in V, \quad \forall a, b \in K. \tag{2}$$

Let us discuss a few examples of linear and nonlinear transformations.

- *Example 1:* The transformation

$$\begin{aligned} F : \mathbb{R} &\to \mathbb{R} \\ x &\to \sin(x) \end{aligned}$$

  is nonlinear. In fact, $\sin(x+y) \neq \sin(x) + \sin(y)$ for arbitrary x and y in $\mathbb{R}$.

- *Example 2:* Let $V = C^{(1)}(\mathbb{R})$ (vector space of real-valued continuously differentiable functions), $W = C^{(0)}(\mathbb{R})$ (vector space of real-valued continuous functions), $K = \mathbb{R}$. The transformation

$$\begin{aligned} F : C^1(\mathbb{R}) &\to C^0(\mathbb{R}) \\ f(x) &\to \frac{df(x)}{dx} \end{aligned}$$

  is linear. In fact, we have

$$\frac{d}{dx}(af(x) + bg(x)) = a\frac{df(x)}{dx} + b\frac{dg(x)}{dx} \qquad \forall f, g \in C^{(1)}(\mathbb{R}), \quad \forall a, b \in \mathbb{R}. \tag{3}$$

- *Example 3:* The transformation

$$\begin{aligned} F : \mathbb{R}^3 &\to \mathbb{R}^2 \\ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &\to \begin{bmatrix} x_1 - x_2 \\ 2x_1 + x_2 - x_3 \end{bmatrix} \end{aligned}$$

  is linear. In fact, we have

$$F\left(a\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + b\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}\right) = \begin{bmatrix} a(x_1 - x_2) + b(y_1 - y_2) \\ a(2x_1 + x_2 - x_3) + b(2y_1 + y_2 - y_3) \end{bmatrix} = aF\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) + bF\left(\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}\right).$$

- *Example 4:* The transformation

$$\begin{aligned} F : \mathbb{R}^3 &\to \mathbb{R}^2 \\ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &\to \begin{bmatrix} x_1 + x_2 + 1 \\ x_3 + x_1 \end{bmatrix} \end{aligned} \tag{4}$$

is <u>*not*</u> linear. In fact,

$$F\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}\right) \neq F\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) + F\left(\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}\right).$$

Transformations of the form (4) are called *affine* transformations. Affine transformations are obtained by adding a constant vector to a linear transformation. For the transformation (4) we have

$$F\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) \to \underbrace{\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\text{linear transformation}} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\text{constant vector}} . \tag{5}$$

- *Example 5:* The transformation[1]

$$\begin{aligned} \text{trace} : M_{n\times n}(\mathbb{R}) &\to \mathbb{R} \\ A &\to \sum_{k=1}^{n} a_{kk} \quad \text{(trace of the matrix } A) \end{aligned} \tag{6}$$

is linear. In fact,

$$\text{trace}(aA + bB) = a\,\text{trace}(A) + b\,\text{trace}(B). \tag{7}$$

Hereafter we show that the composition of two linear transformation is a linear transformation.

Theorem 1. Let U, V, and W be vector spaces. Consider the linear transformations $F : U \to V$ and $G : V \to W$. Then $G(F(u)) : U \to W$ is a linear transformation.

Proof. If F and G are linear transformations then

$$G(F(au + bv)) = G(aF(u) + bF(v)) = aG(F(u)) + bG(F(v)). \tag{8}$$

Hence, the composition of F and G is a linear transformation.

□

[1]The trace of a square matrix is defined to be the sum of all diagonal entries of A.

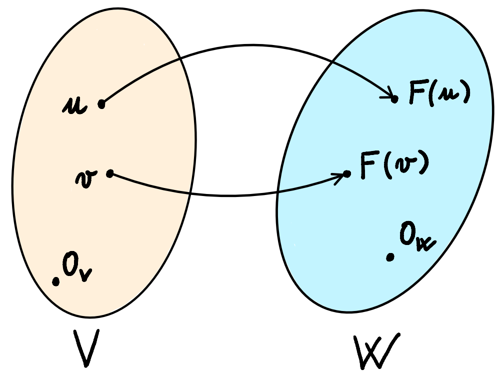
Injective, surjective and invertible transformations. Let V and W be two vector spaces. Consider the following transformation

$$F : V \to W \tag{9}$$

Here, F can be linear on nonlinear.

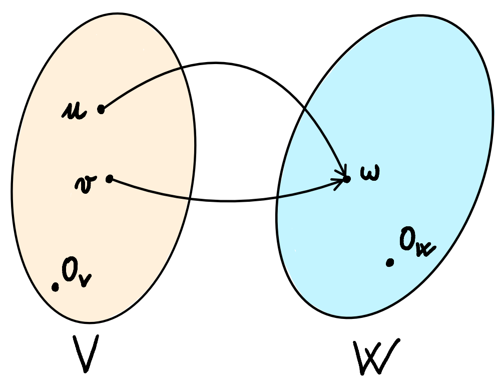
1. We say that F is *injective* or *one-to-one* if:

$$\text{for all} \quad u, v \in V \quad F(u) = F(v) \quad \Rightarrow \quad u = v \tag{10}$$



2. We say that F is *surjective* or *onto* if

$$\text{for all} \quad w \in W \quad \text{there exists (at least one)} \quad u \in V \quad \text{such that} \quad F(u) = w \tag{11}$$



Note that there may be more than one element in V that is mapped onto w. In the figure above, two elements u and v are mapped onto the same element w.

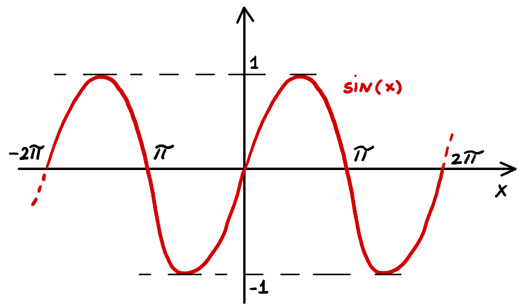
3. We say that F is *invertible*[2] if is is one-to-one and onto (injective and surjective).

[2] Invertible transformations are often called bijections or bijective transformations.

Example 6: The nonlinear transformation

$$\begin{aligned} F : \mathbb{R} &\to \mathbb{R} \\ x &\to \sin(x) \end{aligned}$$

is not injective nor surjective on the real line.



In fact, there are multiple points on the x axis with the same value of $\sin(x)$. For example,

$$\sin(1) = \sin(1 + 2k\pi) \quad k \in \mathbb{Z}. \tag{12}$$

Hence the function is not injective. The function $\sin(x)$ is also not surjective in $\mathbb{R}$, as there is no $x \in \mathbb{R}$ such that $\sin(x) = 2$. However, if we restrict the domain and range of F as follows

$$\begin{aligned} F : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] &\to [-1, 1] \\ x &\to \sin(x) \end{aligned}$$

then F is invertible, since it is injective and surjective. The inverse function is denoted by $\sin^{-1}(x)$ or $\arcsin(x)$

Example 7: The linear transformation

$$\begin{aligned} F : \mathbb{R}^2 &\to \mathbb{R}^2 \\ \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} &\to \underbrace{\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \begin{bmatrix} x_1 + 2x_2 \\ -x_1 + x_2 \end{bmatrix}, \end{aligned}$$

is one-to-one and onto. In fact it is easy to show $Ax = Ay$ implies $x = y$ (inejctivity), and that for each $y \in \mathbb{R}$ there exits $x \in \mathbb{R}^2$ such that $Ax = y$. Therefore the transformation F is invertible. The inverse transformation is defined by the inverse matrix A^{-1}

$$\begin{aligned} F^{-1} : \mathbb{R}^2 &\to \mathbb{R}^2 \\ \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} &\to \underbrace{\frac{1}{3}\begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \frac{1}{3}\begin{bmatrix} x_1 - 2x_2 \\ x_1 + x_2 \end{bmatrix}. \end{aligned}$$

Definition. Let V, W be vector spaces, $F : V \to W$ a linear transformation. If F is invertible then we say that F is an *isomorphism* between V and W. If there exists an isomorphism between the vector spaces V and W (i.e., an invertible linear trasformation) then we say that V and W are *isomorphic*.

Theorem 2. Let V be a vector space of dimension n over a field K. Then V is isomorphic to K^n.

Proof. Let $v_1, \dots, v_n \in V$ be a basis of F. Any vector $v \in V$ can be represented uniquely relative to the basis as

$$v = x_1 v_1 + \cdots + x_n v_n \qquad x_i \in K. \tag{13}$$

The transformation

$$F : V \to K^n \tag{14}$$

$$v \to \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{15}$$

is linear, one-to-one and onto. These properties follow immediately from the definition of basis (surjectivity), and from the fact that the coordinates of $v \in V$ relative to a basis are unique (injectivity). Hence, (15) defines a bijection between V and K^n. This means that V is isomorphic to K^n.

□

Example 8: The space of polynomials of degree at most 4 with real coefficients, i.e., $\mathbb{P}_4(\mathbb{R})$, is isomorphic to $\mathbb{R}^5$. In fact, if we set up a basis for $\mathbb{P}_4(\mathbb{R})$, i.e., a set of 5 linearly independent polynomials of degree at most 4, e.g.,

$$p_4(x) = x^4 - 3x, \quad p_3(x) = x^3, \quad p_2(x) = x^3 + x^2 + 1, \quad p_1(x) = x - x^3, \quad p_0(x) = x^2 + 1, \tag{16}$$

then we see that each polynomial in $p \in \mathbb{P}_4(\mathbb{R})$ is uniquely identified by 5 real coefficients $(x_0, \dots, x_4)$:

$$p(x) = x_4 p_4(x) + x_3 p_3(x) + x_2 p_2(x) + x_1 p_1(x) + x_0 p_0(x). \tag{17}$$

Hence, there exists a bijection between $\mathbb{R}^5$ and the space of polynomials $\mathbb{P}_4(\mathbb{R})$. In other words, $\mathbb{P}_4(\mathbb{R})$ and $\mathbb{R}^5$ are isomorphic.

Example 9: The vector space of 3×3 *symmetric* matrices with real coefficient is isomorphic to $\mathbb{R}^6$.

Since the inverse of an isomorphism is an isomorphism we have that all vector spaces of dimension n over some field K are isomorphic to one another. For example, the vector space of polynomials of degree at most 3 is isomorphic to the vector space of 2×2 matrices with real coefficients.

Theorem 3. The set of all linear mappings between two vector spaces V and W is a vector space. Such a space is denoted by $\mathcal{L}(V, W)$.

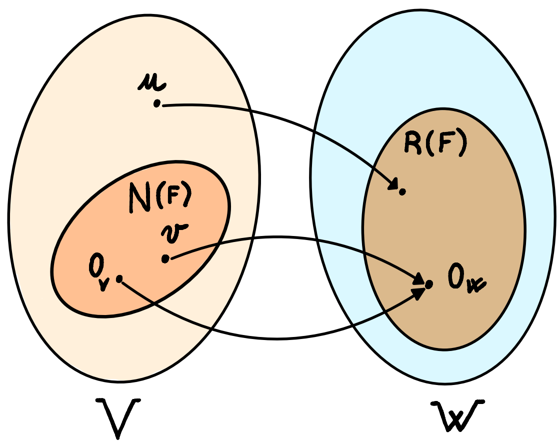
Nullspace and range of a linear transformation. Let V, W be vector spaces. Consider the linear transformation

$$F : V \to W. \tag{18}$$

- The *nullspace* (or kernel[3]) of F is the set vectors in V that are mapped into 0_W (zero vector of W), i.e.,

$$N(F) = \{v \in V \quad \text{such that} \quad F(v) = 0_W\} \qquad \text{(nullspace of } F\text{)}. \tag{19}$$

  Clearly, since F is linear we have that the element 0_V is always mapped onto 0_W. Therefore, 0_V is always in the nullspace of F.



- The *range* of F is the set of vectors w in W such that w is the image of some $v \in V$ under F, i.e., there exists $v \in V$ such that $F(v) = w$.

$$R(F) = \{F(v) \in W \quad \text{such that} \quad v \in V\} \tag{20}$$

  Note that the range of $R(F)$ has 0_W in it. In fact, since F is linear we have that $F(0_V) = 0_W$.

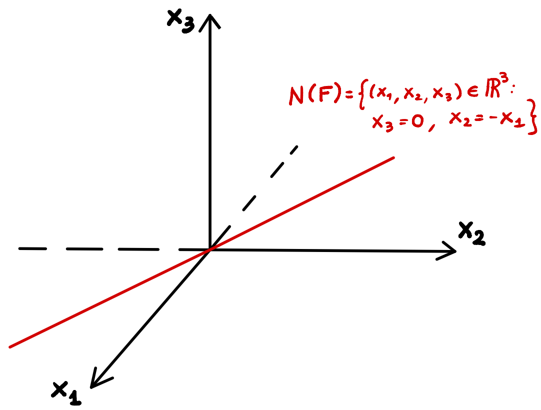
Let us determine the nullspace and the range of simple linear transformations.

Example 10: Consider the following linear transformation

$$\begin{aligned} &F : \mathbb{R}^3 \to \mathbb{R}^2 \\ &\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \to \begin{bmatrix} x_1 + x_2 + x_3 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned} \tag{21}$$

The nullspace of F is the set of vectors in $\mathbb{R}^3$ that mapped onto the zero vector of $\mathbb{R}^2$. Hence, the nullspace of F is defined by the following homogeneous linear system of equations

$$\begin{cases} x_1 + x_2 + x_3 = 0 \\ x_3 = 0 \end{cases} \qquad \Rightarrow \qquad \begin{cases} x_1 = -x_2 \\ x_3 = 0 \end{cases} \tag{22}$$



Note that the nullspace of F is a vector subspace of $\mathbb{R}^3$ (line passing through the origin). The range of F can be constructed by taking an arbitrary element of $\mathbb{R}^3$ and mapping it via F. Such range coincides with *column space* of the matrix A, i.e., the span of the columns of A. In fact,

$$\begin{bmatrix} x_1 + x_2 + x_3 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{23}$$

Hence,

$$R(F) = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} = \mathbb{R}^2. \tag{24}$$

Theorem 4. Let V, W be vector spaces, $F : V \to W$ linear. Then

1. $N(F)$ is a vector subspace of V.
2. $R(F)$ is a vector subspace of W.

Proof. Let $u, v \in N(F)$. Clearly, $u + v$ is in $N(F)$. In fact, since F is linear we have $F(u + v) = F(u) + F(v) = 0_W$. Thus, $u + v$ is in $N(F)$. Moreover, $0_V \in N(F)$ and $cu \in N(F)$ for all $u \in N(F)$ and all $c \in K$. This implies that $N(F)$ is a vector subspace of V. To prove that $R(F)$ is a vector subspace of W, let $w, s \in R(F)$. This means that there exist $u, v \in V$ such that $F(u) = w$ and $F(v) = s$. Obviously, $(w + s) \in R(F)$. In fact, by using the linearity of F we have $F(u + v) = w + s$, and therefore $w + s \in R(F)$. Also, 0_W is $R(F)$ and $cu \in R(F)$ for all $u \in R(F)$. Thus, $R(F)$ is a vector subspace of W.

□

The nullspace and the range of linear transformation also characterize the injectivity and surjectivity of the transformation. In particular we have the following theorems.

Theorem 5. Let V, W be vector spaces, $F : V \to W$ a linear transformation. Then F is injective (one-to-one) if and only if $N(F) = \{0_V\}$, i.e., the if nullspace of F reduces the single element $\{0_V\}$.

[3] The nullspace/kernel of a linear transformation F is often denoted as $\ker(F)$.

Proof. To prove the theorem we need to prove two statements:

1. F is injective $\Rightarrow N(F) = \{0_V\}$.

   Suppose that F is one-to-one. We want to show that this implies $N(F) = \{0_V\}$. To this end, let $v \in N(F)$, i.e., $F(v) = 0_W$. Clearly $v = 0_V$ is mapped onto 0_W, i.e., $0_V \in N(F)$. The assumption that F is one-to-one rules out the existence of any other element in V mapped onto 0_W. In other words, 0_V is the only element of V mapped into 0_W. Hence, if F is one-to-one then $N(F) = \{0_V\}$.

2. $N(F) = \{0_V\} \Rightarrow F$ is injective.

   Conversely, let us assume that $N(F) = \{0_V\}$. We want to show that this implies that F is one-to-one. To this end, suppose there are two elements $u, v \in V$ such that $F(u) = F(v)$. By using the linearity of F we have $F(u - v) = 0_W$, i.e., $(u - v) \in N(F)$. Since, by assumption, the only element in the nullspace of F is 0_V we have that $u - v = 0_V$, i.e., $u = v$. In other words, $N(F) = \{0_V\}$ implies that F is one-to-one.

□

Theorem 6. Let V, W be vector spaces, $F : V \to W$ linear. Then F is surjective (onto) if and only if $\dim(R(F)) = \dim(W)$.

Proof. As before, to prove the theorem we need to prove two statements:

1. F is surjective $\Rightarrow \dim(R(F)) = \dim(W)$,
2. F is surjective $\Leftarrow \dim(R(F)) = \dim(W)$.

Let F be surjective (or onto), i.e., $\forall w \in W$ there exists at least one $v \in V$ such that $F(v) = w$. This means that $R(F) = W$ and therefore $\dim(R(F)) = \dim(W)$. Conversely, suppose that $\dim(R(F)) = \dim(W)$. We know that $R(F)$ is a vector subspace of W. Since the dimension of $R(F)$ and W are the same (by assumption) then $R(F) = W$, i.e., F is surjective (or onto).

□

Next we discuss a very important theorem for linear transformations between vector spaces.

Theorem 7. Let V and W be vector space and $F : V \to W$ be any linear transformation. Then

$$\dim(V) = \dim(N(F)) + \dim(R(F)). \tag{25}$$

Proof. If $R(F) = 0_W$ the statement is trivial since the entire V is mapped to the 0_W. This implies $N(F) = V$, and of course $\dim(N(F)) = \dim(V)$. Consider now $\dim(R(F)) = s > 0$ and let $\{w_1, \ldots, w_s\}$ be a basis of $R(F)$. Then there exist s elements $v_1, \ldots, v_s \in V$ such that $F(v_1) = w_1, \ldots, F(v_s) = w_s$. Suppose $\dim(N(F)) = q$ and let $\{u_1, \ldots, u_q\}$ be a basis for $N(F)$.

We would like to show that $\{u_1, \ldots, u_q, v_1, \ldots, v_s\}$ is a basis of V[4]. To this end, pick an arbitrary $v \in V$. Then, there exists $x_1, \ldots, x_s \in K$ such that $F(v) = x_1 w_1 + \ldots + x_s w_s$ (since $w_1, \ldots, w_s$ is a

[4]Note that if $\{u_1, \ldots, u_q, v_1, \ldots, v_s\}$ is a basis of V then $\dim(V) = q+s$, where $q = \dim(N(F)$ and $s = \dim(R(F))$.

basis for $R(F)$). Recalling that $F(v_1) = w_1$, ..., $F(v_s) = w_s$

$$\begin{aligned} F(v) &= x_1 F(v_1) + \ldots + x_s F(v_s) \\ &= F(x_1 v_1 + \ldots + x_s v_s). \end{aligned}$$

By using the linearity of F we obtain

$$F(v - x_1 v_1 - \cdots - x_s v_s) = 0_W \quad \Rightarrow \quad (v - x_1 v_1 - \cdots - x_s v_s) \in N(F).$$

At this point we represent $(v - x_1 v_1 - \cdots - x_s v_s)$ relative to the basis of $N(F)$

$$v - x_1 v_1 - \ldots - x_s v_s = y_1 u_1 + \ldots + y_q u_q,$$

i.e.,

$$v = x_1 v_1 + \ldots + x_s v_s + y_1 u_1 + \ldots + y_q u_q.$$

This shows that $V = \text{span}\{v_1, \ldots, v_s, u_1, \ldots, u_q\}$, i.e., that V is generated by $\{v_1, \ldots, v_s, u_1, \ldots, u_q\}$. To prove the theorem it remains to prove that the the vectors $\{v_1, \ldots, v_s, u_1, \ldots, u_q\}$ are linearly independent. In this way we can claim that $n = s + q$, i.e., $\dim(V) = \dim(N(F)) + \dim(R(F))$.

To this end, consider the linear combination

$$x_1 v_1 + \ldots + x_s v_s + y_1 u_1 + \ldots + y_q u_q = 0_V. \tag{26}$$

By applying F and recalling that $F(u_i) = 0_W$ ($u_i \in N(F)$) we obtain

$$x_1 w_1 + \ldots + x_s w_s = 0_W \quad \Rightarrow \quad x_1, \ldots, x_s = 0. \tag{27}$$

In fact $\{w_1, \ldots, w_s\}$ is a basis for $R(F)$ and therefore w_i are linearly independent. Substituting this result back into (26) yields

$$y_1 u_1 + \ldots + y_q u_q = 0_V \quad \Rightarrow \quad y_1, \ldots, y_q = 0 \tag{28}$$

since $\{u_1, \ldots, u_q\}$ is a basis for $N(F)$. Equations (27), (28) and (26) allow us to conclude that $\{v_1, \ldots, v_s, u_1, \ldots, u_q\}$ are linearly independent. Moreover the vectors $\{v_1, \ldots, v_s, u_1, \ldots, u_q\}$ generate V, and therefore they are a basis for V. This implies that

$$\dim(V) = s + q = \dim(N(F)) + \dim(R(F)). \tag{29}$$

□

Matrix rank theorem. Theorem 7 can be applied to linear transformations defined by matrices. To this end, consider the transformation $F : \mathbb{R}^n \to \mathbb{R}^m$ from $\mathbb{R}^n$ into $\mathbb{R}^m$ defined as $F(x) = Ax$, where A is an $m \times n$ matrix:

$$\underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{x} \to \underbrace{\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{x} \tag{30}$$

We know that range of F coincides with the column space of A. Also the dimension of the column space is the *rank* of the matrix A. Therefore from equation (25) it follows that

$$\boxed{n = \dim(N(A)) + \text{rank}(A)}. \tag{31}$$

Matrix associated with a linear transformation Let V and W be finite-dimensional vector spaces, and let

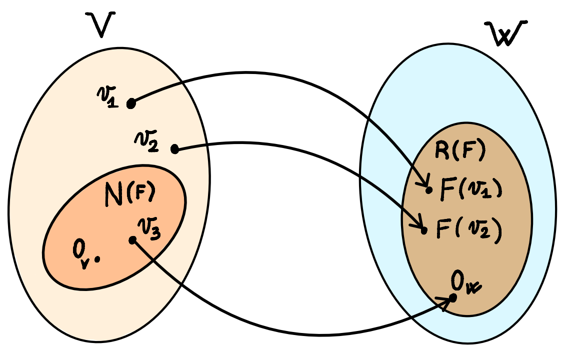
$$F : V \to W \tag{32}$$

an arbitrary linear transformation. In this section we show how to represent F in terms of a matrix. To this end, suppose that

$$\begin{aligned} \mathcal{B}_V &= \{v_1, \ldots, v_n\} & \to & \quad \text{basis of } V, \quad \dim(V) = n, \\ \mathcal{B}_W &= \{w_1, \ldots, w_m\} & \to & \quad \text{basis of } W, \quad \dim(W) = m. \end{aligned}$$

The transformation F is *uniquely determined* by the image of the basis $\mathcal{B}_V$ under F, i.e.,

$$\{v_1, \ldots, v_n\} \quad \to \quad \{F(v_1), \ldots, F(v_n)\}. \tag{33}$$



Clearly, for all $i = 1, \ldots, n$ we have that $F(v_i) \in R(F) \subseteq W$. Therefore, each $F(v_i)$ can be represented in terms of the basis $\mathcal{B}_W$ as

$$\begin{cases} F(v_1) = a_{11}w_1 + \cdots + a_{m1}w_m \\ \quad \vdots \\ F(v_n) = a_{1n}w_1 + \cdots + a_{mn}w_m \end{cases}. \tag{34}$$

Note that a_{ij} is the i-th component of $F(v_j)$ relative to the basis $\{w_1, \ldots, w_m\}$. The matrix associated with the linear transformation F depends bases $\mathcal{B}_V$ and $\mathcal{B}_W$ and it is defined as

$$A_{\mathcal{B}_V}^{\mathcal{B}_W}(F) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}. \tag{35}$$

Next, consider an arbitrary element $v \in V$, and represent it in terms of the basis $\mathcal{B}_V$

$$v = x_1v_1 + \cdots + x_nv_n. \tag{36}$$

By applying F and taking (34) into account we obtain

$$\begin{aligned} F(v) =& x_1 F(v_1) + \cdots + x_n F(v_n) \\ =& x_1 \left(a_{11} w_1 + \cdots + a_{m1} w_m\right) + \cdots + x_n \left(a_{1n} w_1 + \cdots + a_{mn} w_m\right) \\ =& \underbrace{\left(a_{11} x_1 + \cdots + a_{1n} x_n\right)}_{y_1} w_1 + \cdots + \underbrace{\left(a_{m1} x_1 + \cdots + a_{mn} x_n\right)}_{y_m} w_m. \end{aligned} \tag{37}$$

At this point we define the following two column vectors

$$[v]_{\mathcal{B}_V} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad [F(v)]_{\mathcal{B}_W} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \tag{38}$$

representing the *coordinates* of v and $F(v)$ relative to the bases $\mathcal{B}_V$ and $\mathcal{B}_W$, respectively[5]. With this notation, we see from (37) and (35) that

$$[F(v)]_{\mathcal{B}_W} = A^{\mathcal{B}_W}_{\mathcal{B}_V}(F)[v]_{\mathcal{B}_V}. \tag{39}$$

Therefore, the coordinates of $F(v)$ relative to $\mathcal{B}_W$ are obtained by taking the matrix-vector product between the matrix $A^{\mathcal{B}_W}_{\mathcal{B}_V}(F)$ and the coordinates of v relative to $\mathcal{B}_V$.

Example 11: Let V and W be vector spaces of dimension $\dim(V) = 2$ and $\dim(W) = 3$, respectively. We consider the following bases in V and W:

$$\mathcal{B}_V = \{v_1, v_2\}, \qquad \mathcal{B}_W = \{w_1, w_2, w_3\}. \tag{40}$$

Relative to such bases, suppose that F is defined as

$$\begin{cases} F(v_1) = w_1 - 2w_2 - w_3 \\ F(v_2) = w_1 + w_2 + w_3 \end{cases}. \tag{41}$$

Then the matrix representing F is

$$A^{\mathcal{B}_W}_{\mathcal{B}_V}(F) = \begin{bmatrix} 1 & 1 \\ -2 & 1 \\ -1 & 1 \end{bmatrix}. \tag{42}$$

If $v = x_1 v_1 + x_2 v_2$ is an arbitrary vector in V then

$$\begin{aligned} F(v) =& x_1 F(v_1) + x_2 F(v_2) \\ =& x_1(w_1 - 2w_2 - w_3) + x_2(w_1 + w_2 + w_3) \\ =& \underbrace{(x_1 + x_2)}_{y_1} w_1 + \underbrace{(x_2 - 2x_1)}_{y_2} w_2 + \underbrace{(x_2 - x_1)}_{y_3} w_3. \end{aligned} \tag{43}$$

Note that the coordinates of $F(v)$ relative to the basis $\mathcal{B}_W$, i.e., $\{y_1, y_2, y_3\}$ are given by the standard matrix-vector product

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{44}$$

[5]We know from Lecture 6 that such coordinates are uniquely defined by the basis.

Change of basis transformation Consider the following two bases in the vector space V

$$\mathcal{B}_1 = \{u_1, \ldots, u_n\}$$
$$\mathcal{B}_2 = \{v_1, \ldots, v_n\}$$

Obviously, we can express any element in $\mathcal{B}_1$ as a linear combination of elements in $\mathcal{B}_2$ and vice versa. For example,

$$\begin{cases} v_1 = \alpha_{11}u_1 + \cdots + \alpha_{n1}u_m \\ \quad\vdots \\ v_n = \alpha_{1n}u_1 + \cdots + \alpha_{nn}u_n \end{cases}. \tag{45}$$

The matrix associated with the linear transformation "change of basis *from* $\mathcal{B}_2$ *to* $\mathcal{B}_1$" is

$$M_{\mathcal{B}_2}^{\mathcal{B}_1} = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{bmatrix}. \tag{46}$$

Such a matrix is invertible and it allows us to transform the coordinates of any vector $v \in V$ from those relative to $\mathcal{B}_1$ to those relative to $\mathcal{B}_2$, i.e.,

$$[v]_{\mathcal{B}_2} = M_{\mathcal{B}_1}^{\mathcal{B}_2}[v]_{\mathcal{B}_1}. \tag{47}$$

Moreover, we have

$$[v]_{\mathcal{B}_1} = M_{\mathcal{B}_2}^{\mathcal{B}_1}[v]_{\mathcal{B}_2} = \left(M_{\mathcal{B}_1}^{\mathcal{B}_2}\right)^{-1}[v]_{\mathcal{B}_2} \quad \text{which implies} \quad M_{\mathcal{B}_2}^{\mathcal{B}_1} = \left(M_{\mathcal{B}_1}^{\mathcal{B}_2}\right)^{-1}. \tag{48}$$

The change of basis transformation can be also used to represent a linear transformation $F : V \to W$ relative to different bases in V and W. To show this, let

$$\begin{aligned} \mathcal{B}_1, \mathcal{B}_2 &\quad\to\quad \text{Bases of } V, \quad \dim(V) = n, \\ \mathcal{B}_3, \mathcal{B}_4 &\quad\to\quad \text{Bases of } W, \quad \dim(W) = m. \end{aligned}$$

We have,

$$[F(v)]_{\mathcal{B}_4} = M_{\mathcal{B}_3}^{\mathcal{B}_4}[F(v)]_{\mathcal{B}_3} = M_{\mathcal{B}_3}^{\mathcal{B}_4}A_{\mathcal{B}_2}^{\mathcal{B}_3}[v]_{\mathcal{B}_2} = \underbrace{M_{\mathcal{B}_3}^{\mathcal{B}_4}A_{\mathcal{B}_2}^{\mathcal{B}_3}M_{\mathcal{B}_1}^{\mathcal{B}_2}}_{A_{\mathcal{B}_1}^{\mathcal{B}_4}}[v]_{\mathcal{B}_1}, \tag{49}$$

i.e.,

$$A_{\mathcal{B}_1}^{\mathcal{B}_4} = M_{\mathcal{B}_3}^{\mathcal{B}_4}A_{\mathcal{B}_2}^{\mathcal{B}_3}M_{\mathcal{B}_1}^{\mathcal{B}_2}. \tag{50}$$

The matrix $A_{\mathcal{B}_1}^{\mathcal{B}_4}$ represents the linear transformation F relative to the bases $\mathcal{B}_1$ (basis of V) and $\mathcal{B}_4$ (basis of W). Similarly, $A_{\mathcal{B}_2}^{\mathcal{B}_3}$ represents the linear transformation F relative to the bases $\mathcal{B}_2$ (basis of V) and $\mathcal{B}_3$ (basis of W).

Example 12: (Change of basis in $\mathbb{R}^2$) Consider the following bases of $\mathbb{R}^2$

$$\mathcal{B}_1 = \{e_1, e_2\}, \qquad e_1 = \begin{bmatrix}1\\0\end{bmatrix}, \quad e_2 = \begin{bmatrix}0\\1\end{bmatrix} \quad \text{(canonical basis of } \mathbb{R}^2\text{)},$$
$$\mathcal{B}_2 = \{v_1, v_2\}, \qquad v_1 = \begin{bmatrix}1\\1\end{bmatrix}, \quad v_2 = \begin{bmatrix}1\\2\end{bmatrix}.$$

Define the change of basis transformation

$$F : \mathbb{R}^2 \to \mathbb{R}^2 \tag{51}$$

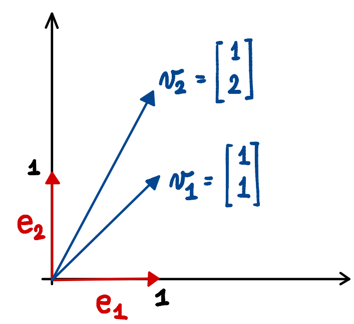
as

$$F\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}1\\1\end{bmatrix}, \qquad F\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}1\\2\end{bmatrix}. \tag{52}$$

Clearly,

$$\begin{cases} v_1 = e_1 + e_2 \\ v_2 = e_1 + 2e_2 \end{cases}. \tag{53}$$

The following figure sketches $\{e_1, e_2\}$ and $\{v_1, v_2\}$ as vectors in the Cartesian plane.



Any vector $v \in \mathbb{R}^2$ can be expressed relatively to $\mathcal{B}_1$ or $\mathcal{B}_2$:

$$\begin{aligned} v =& x_1 v_1 + x_2 v_2 \\ =& x_1(e_1 + e_2) + x_2(e_1 + 2e_2) \\ =& (x_1 + x_2)e_1 + (x_1 + 2x_2)e_2. \end{aligned} \tag{54}$$

Denote by

$$[v]_{\mathcal{B}_1} = \begin{bmatrix}y_1\\y_2\end{bmatrix}, \qquad [v]_{\mathcal{B}_2} = \begin{bmatrix}x_1\\x_2\end{bmatrix}. \tag{55}$$

the coordinates of v relative to $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. Then equation (54) implies that

$$[v]_{\mathcal{B}_1} = M_{\mathcal{B}_2}^{\mathcal{B}_1}[v]_{\mathcal{B}_2}, \qquad \text{where} \qquad M_{\mathcal{B}_2}^{\mathcal{B}_1} = \begin{bmatrix}1 & 1\\1 & 2\end{bmatrix}. \tag{56}$$

$M_{\mathcal{B}_2}^{\mathcal{B}_1}$ is the matrix associated with the change of basis transformation $\mathcal{B}_2 \to \mathcal{B}_1$. Clearly, $M_{\mathcal{B}_2}^{\mathcal{B}_1}$ is invertible with inverse

$$M_{\mathcal{B}_1}^{\mathcal{B}_2} = \left(M_{\mathcal{B}_2}^{\mathcal{B}_1}\right)^{-1} = \begin{bmatrix}2 & -1\\-1 & 1\end{bmatrix} \tag{57}$$

$M_{\mathcal{B}_1}^{\mathcal{B}_2}$ is the matrix associated with the change of basis transformation $\mathcal{B}_1 \to \mathcal{B}_2$. Let us see if this is true. To this end, we consider the vector $v = e_1$ and compute the coordinates of this vector relative to $\mathcal{B}_2$. We have

$$[v]_{\mathcal{B}_1} = \begin{bmatrix}1\\0\end{bmatrix} \quad \Rightarrow \quad [v]_{\mathcal{B}_2} = \underbrace{\begin{bmatrix}2 & -1\\-1 & 1\end{bmatrix}}_{M_{\mathcal{B}_1}^{\mathcal{B}_2}} \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}2\\-1\end{bmatrix} \tag{58}$$

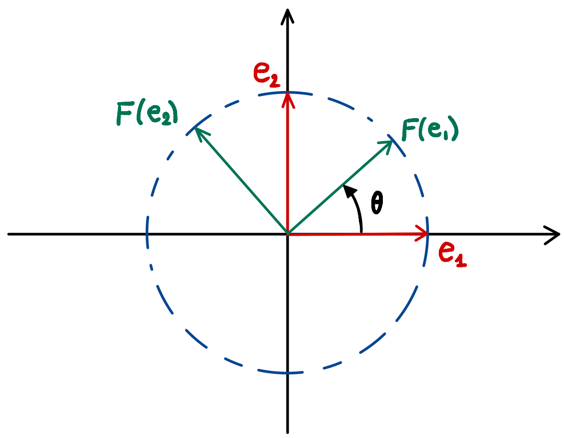
Example 13: (Rotations in $\mathbb{R}^2$) Consider the linear transformation $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined as follows (counterclockwise rotation of the basis vectors by an angle θ)

$$\begin{cases} F(e_1) = \cos(\theta)e_1 + \sin(\theta)e_2 \\ F(e_2) = -\sin(\theta)e_1 + \cos(\theta)e_2 \end{cases}, \tag{59}$$

where

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{60}$$

is the canonical basis of $\mathbb{R}^2$.



The matrix associated with the transformation F relative to the basis $\mathcal{B}_V = \{e_1, e_2\}$ is

$$A^{\mathcal{B}_V}_{\mathcal{B}_V}(F) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \qquad \text{(2D rotation matrix).} \tag{61}$$

Any vector with components $[v]_{\mathcal{B}_V}$ is rotated to a vector $F(v)$ with components

$$[F(v)]_{\mathcal{B}_V} = A^{\mathcal{B}_V}_{\mathcal{B}_V}[v]_{\mathcal{B}_V}. \tag{62}$$

For example, the vector $v = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ has components $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ relative to the canonical basis $\mathcal{B}_V$, and it is transformed to a vector $F(v)$ with components

$$[F(v)]_{\mathcal{B}_V} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2\cos(\theta) - \sin(\theta) \\ 2\sin(\theta) + \cos(\theta) \end{bmatrix}. \tag{63}$$

In particular, if $\theta = \pi/2$ (90 degrees counterclockwise rotation) then

$$[F(v)]_{\mathcal{B}_V} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}. \tag{64}$$

The inverse transformation (inverse rotation) is obtained by replacing θ with $-\theta$ in (61), i.e.,

$$\left[A^{\mathcal{B}_V}_{\mathcal{B}_V}(F)\right]^{-1} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}. \tag{65}$$

It is straightforward to verify that for all $\theta \in [0, 2\pi]$ we have

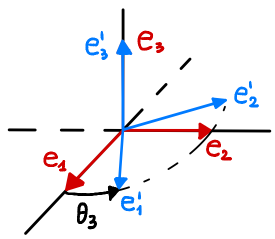
$$\left[A_{\mathcal{B}_V}^{\mathcal{B}_V}(F)\right]^{-1} A_{\mathcal{B}_V}^{\mathcal{B}_V}(F) = I_2. \tag{66}$$

In fact,

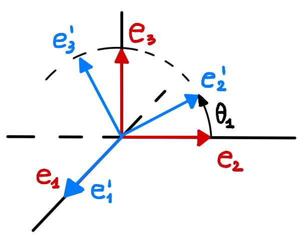
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The rotation matrix is an *orthogonal matrix*[6].

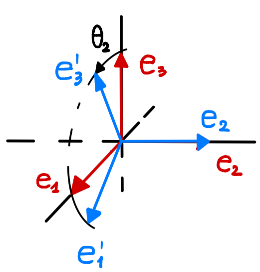
Example 14: (Rotations in $\mathbb{R}^3$) We can define rotations along each of the three axes of a 3D Cartesian coordinate system, i.e.,



$$R_3 = \begin{bmatrix} \cos\theta_3 & -\sin\theta_3 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



$$R_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{bmatrix}$$



$$R_2 = \begin{bmatrix} \cos\theta_2 & 0 & -\sin\theta_2 \\ 0 & 1 & 0 \\ \sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix}$$

Note that the composition of two rotations in $\mathbb{R}^3$ does not commute. For example,

$$R_1 R_3 \neq R_3 R_1.$$

[6]In general, we say that $A \in M_{n\times n}(\mathbb{R})$ is orthogonal if

$$A^T = A^{-1}. \tag{67}$$

This is equivalent to the statement that orthogonal matrices satisfy

$$AA^T = I_n. \tag{68}$$

Example 15: (Orthogonal projection) Consider

$$F : \mathbb{R}^3 \to \mathbb{R}^3$$

and the canonical bases of $\mathbb{R}^3$

$$\mathcal{B}_3 = \{e_1, e_2, e_3\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

We define F by mapping the basis $\mathcal{B}_3$ as follows

$$F(e_1) = e_1, \qquad F(e_2) = e_2, \qquad F(e_3) = 0_{\mathbb{R}^3}.$$

The associated matrix defines an orthogonal projection onto the (x_1, x_2)

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{69}$$

Note that $P^2 = P$. The orthogonal projection transformation basically project any vector $v \in \mathbb{R}^3$ onto the plane spanned by e_1 and e_2. If we are interested in a projection onto different plane, we can use e.g., the 3D rotation matrices R_i and rotate the plane before applying the projection. Note that with just R_1 and R_3 we can orient the plane (x_1, x_2) in all possible directions. We maintain that

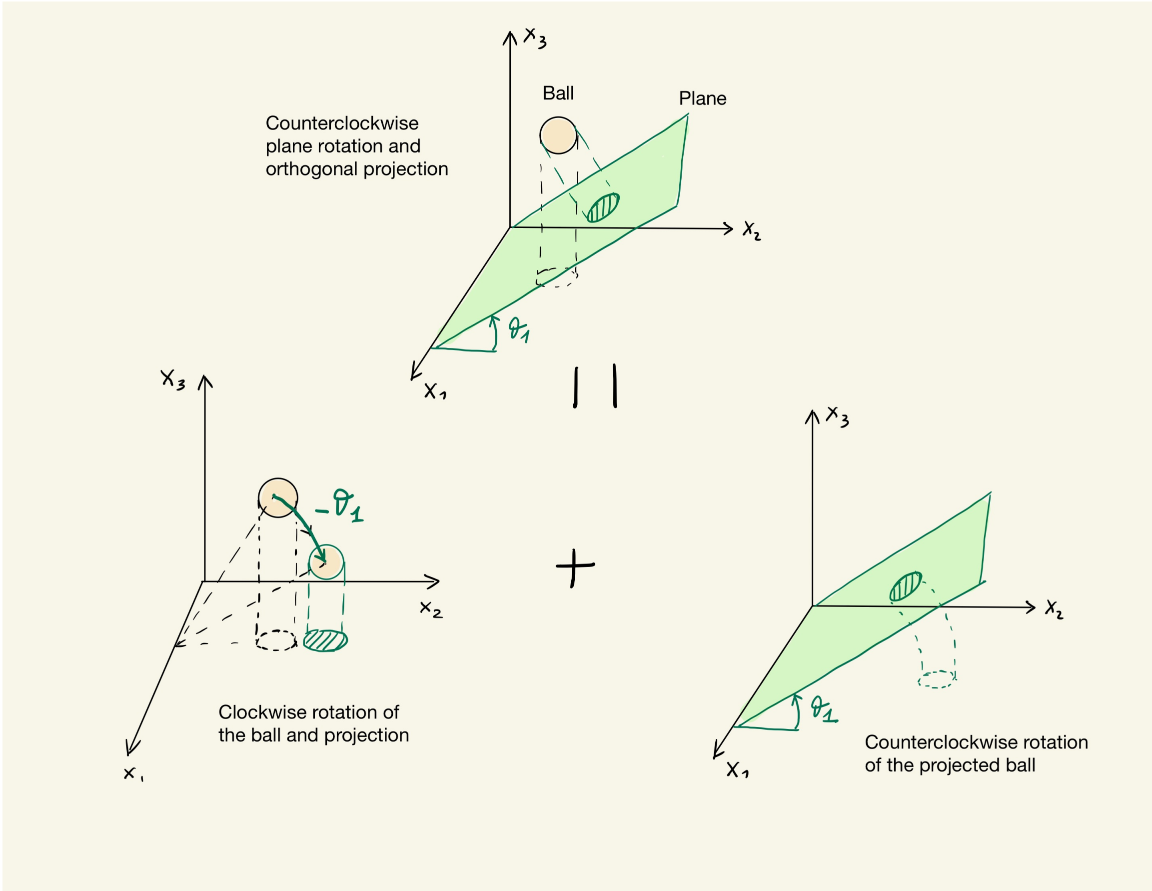
$$P(\theta_1, \theta_3) = R_1(\theta_1) R_3(\theta_3) P R_3^T(\theta_3) R_1^T(\theta_1) \tag{70}$$

is an orthogonal projection onto a tilted plane identified by the angles (θ_1, θ_3). To explain this formula suppose for simplicity that we just rotate the plane (x_1, x_2) counterclockwise of an angle θ_1 around the x_1 axis. The projection of any object onto such plane is obtained by rotating the object clockwise of an angle θ_1 around x_1 (matrix $R_1^T(\theta_3)$ projecting onto the (x_1, x_2) plane and then rotating the result back (matrix $R_1(\theta_3)$). Clearly, (70) satisfies the condition for orthogonal projections,

$$P^2(\theta_1, \theta_3) = P(\theta_1, \theta_3). \tag{71}$$

Example 16: (Oblique projection) Let $v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ be a vector of $\mathbb{R}^3$ representing the direction of a light beam. A light beam passing through an arbitrary point $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ has the form

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = c \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \text{where } c \in \mathbb{R} \tag{72}$$



If we set $y_3 = 0$ we obtain $c = -x_3/v_3$. With such a value for c, the light beam passing through the point x intersects the horizontal plane. The linear transformation defined by

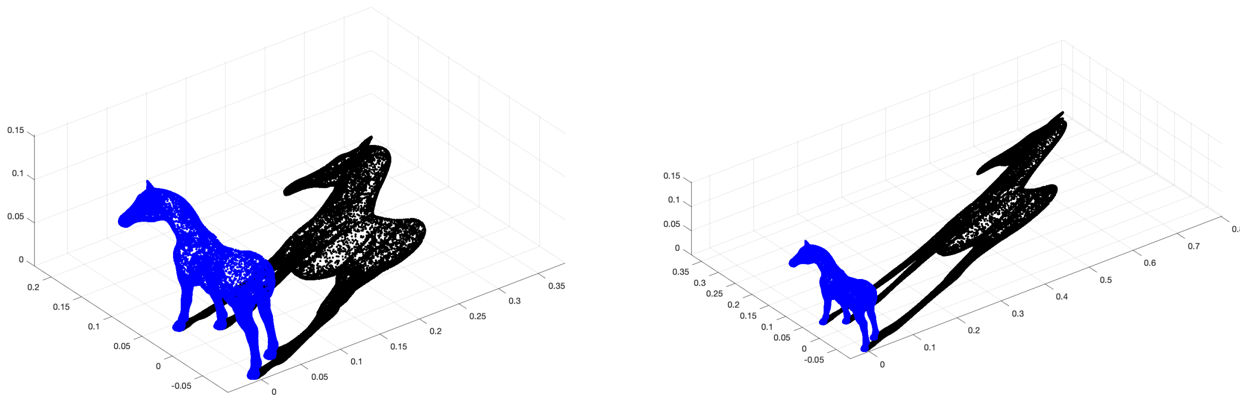
$$\begin{cases} y_1 = -\dfrac{v_1}{v_3} x_3 + x_1 \\ y_2 = -\dfrac{v_2}{v_3} x_3 + x_2 \\ y_3 = 0 \end{cases} \tag{73}$$

defines an oblique projection onto the horizontal plane. The matrix associated with such oblique projection transformation (relative to the canonical basis of $\mathbb{R}^3$) is

$$P = \begin{bmatrix} 1 & 0 & -v_1/v_3 \\ 0 & 1 & -v_2/v_3 \\ 0 & 0 & 0 \end{bmatrix} \tag{74}$$

The oblique projection can be used to compute the *shadow* of any object in 3D. The following figure shows the shadow projected by a horse for various angles of the light beam.

Note that for $v_1 = v_2 = 0$ the oblique projection reduces to the projection we studied in the previous example.



Example 17: Let $\mathbb{P}_4 = \text{span}\{1, x, x^2, x^3, x^4\}$ be the space of polynomials of degree at most 4. Define the linear transformation

$$\begin{aligned} F: \quad & \mathbb{P}_4 \to \mathbb{P}_3 \\ & p(x) \to \frac{dp(x)}{dx}. \end{aligned}$$

The canonical bases of $\mathbb{P}_4$ and $\mathbb{P}_3$ are

$$\begin{aligned} \mathcal{B}_4 &= \{1, x, x^2, x^3, x^4\}, \\ \mathcal{B}_3 &= \{1, x, x^2, x^3\}. \end{aligned}$$

We define the derivative transformation by mapping each element of $\mathbb{P}_4$ and representing the result in terms of $\mathbb{P}_3$. This yields

$$F(1) = 0, \qquad F(x) = 1, \qquad F(x^2) = 2x, \qquad F(x^3) = 3x^2, \qquad F(x^4) = 4x^3.$$

The matrix associated with F (derivative operator) relative to the bases $\mathcal{B}_4$ and $\mathcal{B}_3$ is

$$A_{\mathcal{B}_4}^{\mathcal{B}_3} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}.$$

For example, let us compute the derivative of the polynomial

$$p(x) = 1 - 3x + 6x^3. \tag{75}$$

The coordinates of $p(x)$ relative to $\mathcal{B}_4$ are

$$[p(x)]_{\mathcal{B}_4} = \begin{bmatrix} 1 & -3 & 0 & 6 & 0 \end{bmatrix}^T.$$

This implies that

$$\Rightarrow \left[\frac{dp(x)}{dx}\right]_{\mathcal{B}_3} = A_{\mathcal{B}_4}^{\mathcal{B}_3}[p(x)]_{\mathcal{B}_4} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ 0 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 18 \\ 0 \end{bmatrix}.$$

Therefore we obtained

$$\frac{dp(x)}{dx} = -3 + 0x + 18x^2 + 0x^3 = -3 + 18x^2, \tag{76}$$

which is indeed the derivative of the polynomial (75).