Linear transformations and representations

COMS 3251 Fall 2022 (Daniel Hsu)

1 Elementary linear operators

1.1 Linear functionals

A linear transformation $T \colon \mathbb{R}^n \to \mathbb{R}$ with output space $\mathbb{R}$ is called a *linear functional* (on $\mathbb{R}^n$). For every such linear functional T, there is a $1 \times n$ matrix, which corresponds to a row vector $\mathbf{w}^\top$ for some n-vector $\mathbf{w}$, such that the value of $T(\mathbf{v})$ is given by multiplying the matrix $\mathbf{w}^\top$ by the vector $\mathbf{v}$:

$$T(\mathbf{v}) \;=\; \begin{bmatrix} \leftarrow & \mathbf{w}^\top & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ \mathbf{v} \\ \downarrow \end{bmatrix}.$$

We will use $\mathbf{w}^\top$ to refer to both the row vector corresponding to $\mathbf{w} \in \mathbb{R}^n$ and the linear functional that sends an n-vector $\mathbf{v}$ to $\mathbf{w}^\top\mathbf{v}$.

1.2 Hyperplanes

We introduce a "geometric" interpretation of linear functionals. In this interpretation, n-vectors are regarded as *points* in n-dimensional Cartesian space.

If $\mathbf{w} = (w_1, w_2, w_3)$ is a non-zero 3-vector, then the solution set for the homogeneous linear equation (in variables $\mathbf{x}$)

$$\mathbf{w}^\top\mathbf{x} \;=\; 0$$

is a 2-dimensional subspace (by the Dimension Theorem). We can express this solution set as both $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{w}^\top\mathbf{x} = 0\}$ and $\mathsf{NS}(\mathbf{w}^\top)$. A 2-dimensional subspace of $\mathbb{R}^3$ is called a *plane*.

The solution set for

$$\mathbf{w}^\top\mathbf{x} \;=\; b,$$

where b is an arbitrary real number that is possibly non-zero, is given by

$$\{\mathbf{x}^\star\} + \mathsf{NS}(\mathbf{w}^\top) \;=\; \{\mathbf{x}^\star + \mathbf{v} : \mathbf{v} \in \mathsf{NS}(\mathbf{w}^\top)\},$$

where $\mathbf{x}^\star$ is any particular solution to $\mathbf{w}^\intercal\mathbf{x} = b$. This is the set obtained by taking every vector in $\mathsf{NS}(\mathbf{w}^\intercal)$ and adding $\mathbf{x}^\star$ to it. So, we start with a palne and then (possibly) shift it away from the origin. Such a set is called an *affine plane*. The affine planes corresponding to different values of b are all "parallel" to each other; varying b changes how much the affine plane is shifted away from the origin.

Every 3-vector is on exactly one of two sides of an affine plane, or it is on the affine plane itself. So, in this sense, an affine plane splits 3-dimensional Cartesian space into two parts.

Example. The 3-vector $\mathbf{w} = (3, 4, 5)$ defines the following homogeneous linear equation in the unknown variables (x_1, x_2, x_3):

$$3x_1 + 4x_2 + 5x_3 \;=\; 0.$$

The solution set, i.e., the nullspace of $\mathbf{w}^\intercal$, is the plane spanned by the vectors $(-4/3, 1, 0)$ and $(-5/3, 0, 1)$. For example, another vector in this plane is

$$\begin{bmatrix} 1/3 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -4/3 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} -5/3 \\ 0 \\ 1 \end{bmatrix}.$$

Now consider the linear equation

$$3x_1 + 4x_2 + 5x_3 \;=\; 1.$$

A particular solution is $(-1, 1, 0)$, and so the solution set is

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + c_1 \begin{bmatrix} -4/3 \\ 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} -5/3 \\ 0 \\ 1 \end{bmatrix} : (c_1, c_2) \in \mathbb{R}^2 \right\}.$$ ∎

The concept of planes are generalized to arbitrary dimensions (at least 1) by adding the prefix "hyper". If $\mathbf{w}$ is a non-zero n-vector, then the solution set $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\intercal\mathbf{x} = 0\}$ for the homogeneous linear equation $\mathbf{w}^\intercal\mathbf{x} = 0$ is called a *hyperplane*, and it is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. If b is a real number that is possibly non-zero, then the solution set for $\mathbf{w}^\intercal\mathbf{x} = b$ is generally called an *affine hyperplane*.

Caution. Sometimes the term "hyperplane" is used for affine hyperplanes that do not necessarily contain the origin. In these contexts, the term "homogeneous hyperplane" is used to refer to hyperplanes that contain the origin.

1.3 Projections

For any (non-zero) linear functional $\mathbf{w}^\mathsf{T}$ on $\mathbb{R}^n$, and any (non-zero) n-vector $\mathbf{z}$ such that $\mathbf{w}^\mathsf{T}\mathbf{z} \neq 0$, consider the linear operator P defined as follows:

$$P = \frac{1}{\mathbf{w}^\mathsf{T}\mathbf{z}} \mathbf{z}\mathbf{w}^\mathsf{T}.$$

This operator is called the *elementary projection operator* (a.k.a. *elementary projector*) to $\mathrm{span}(\{\mathbf{z}\})$ along $\mathsf{NS}(\mathbf{w}^\mathsf{T})$. Let us interpret its effect on $\mathbf{v} \in \mathbb{R}^n$.

Interpretation 1. We first give an "algebraic" interpretation of $P\mathbf{v}$.

- Recall that $\mathbf{z}\mathbf{w}^\mathsf{T}$ is interpreted as matrix multiplication: an $n \times 1$ matrix $\mathbf{z}$ multiplying a $1 \times n$ matrix $\mathbf{w}^\mathsf{T}$.
- So to multiply $\mathbf{z}\mathbf{w}^\mathsf{T}$ by the vector $\mathbf{v}$, we first multiply the $1 \times n$ matrix $\mathbf{w}^\mathsf{T}$ by $\mathbf{v}$, which results in the 1-vector $\mathbf{w}^\mathsf{T}\mathbf{v}$.
- Then we multiply the $n \times 1$ matrix $\mathbf{z}$ by the 1-vector $\mathbf{w}^\mathsf{T}\mathbf{v}$, which results in the n-vector $(\mathbf{w}^\mathsf{T}\mathbf{v})\mathbf{z}$.
- The elementary projector P also has the leading factor of $1/(\mathbf{w}^\mathsf{T}\mathbf{z})$, so finally, we obtain the following formula for $P\mathbf{v}$:

  $$P\mathbf{v} = \frac{1}{\mathbf{w}^\mathsf{T}\mathbf{z}} (\mathbf{w}^\mathsf{T}\mathbf{v})\,\mathbf{z} = \frac{\mathbf{w}^\mathsf{T}\mathbf{v}}{\mathbf{w}^\mathsf{T}\mathbf{z}}\,\mathbf{z}.$$

  This is a vector in the span of $\mathbf{z}$.

Interpretation 2. Now we interpret $P\mathbf{v}$ "geometrically", as the intersection between an affine hyperplane and a line.

- Consider the affine hyperplane defined by the linear equation

  $$\mathbf{w}^\mathsf{T}\mathbf{x} = \mathbf{w}^\mathsf{T}\mathbf{v}.$$

  It is clear that this affine hyperplane contains $\mathbf{v}$.

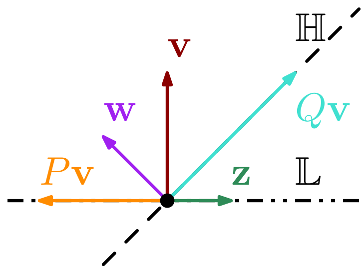

Figure 1: Applying an elementary projector $P = \frac{1}{\mathbf{w}^\mathsf{T}\mathbf{z}}\mathbf{z}\mathbf{w}^\mathsf{T}$, as well as its complement $Q = I - P$, to a vector $\mathbf{v}$.

- Also consider the line in the direction of $\mathbf{z}$, i.e.,

$$\operatorname{span}(\{\mathbf{z}\}) = \{c\,\mathbf{z} : c \in \mathbb{R}\}.$$

- The affine hyperplane and the line intersect at a point $\mathbf{u}$ that satisfies

$$\mathbf{w}^\mathsf{T}\mathbf{u} = \mathbf{w}^\mathsf{T}\mathbf{v} \quad \text{and} \quad \mathbf{u} = c\,\mathbf{z} \quad \text{for some real number } c.$$

  The real number c that makes this work is $c = (\mathbf{w}^\mathsf{T}\mathbf{v})/(\mathbf{w}^\mathsf{T}\mathbf{z})$.

- So the intersection of the affine hyperplane and the line is

$$\mathbf{u} = \frac{\mathbf{w}^\mathsf{T}\mathbf{v}}{\mathbf{w}^\mathsf{T}\mathbf{z}}\mathbf{z},$$

  which is equal to $P\mathbf{v}$.

Example. Consider the Cartesian plane. Let $\mathbf{w} = (-1, 1)$ and $\mathbf{z} = (1, 0)$, and note that $\mathbf{w}^\mathsf{T}\mathbf{z} = -1 \neq 0$. The line is $\mathbb{L} = \{(c, 0) : c \in \mathbb{R}\}$, and the (homogeneous) hyperplane is $\mathbb{H} = \{(x_1, x_2) \in \mathbb{R}^2 : -x_1 + x_2 = 0\}$. (A hyperplane in $\mathbb{R}^2$ is also a line.) Consider the elementary projector $P = -\mathbf{z}\mathbf{w}^\mathsf{T}$. For $\mathbf{v} = (0, 2)$, we have $P\mathbf{v} = -2\mathbf{z} = (-2, 0)$. Note that $\mathbf{v} - P\mathbf{v} \in \mathbb{H}$, so P sends $\mathbf{v}$ to the line $\mathbb{L}$ by adding a vector from $\mathbb{H}$. See Figure 1. ∎

Elementary projection operators are special cases of *projection operators* (a.k.a. *projectors*), which are linear operators that satisfy a property called *idempotency*. We say an operator T is *idempotent* if the composition of the T with itself (i.e., TT, also written T^2) is the same as the operator T itself:

$$TT = T.$$

Let us algebraically verify that the elementary projector $P = \frac{1}{\mathbf{w}^\mathsf{T}\mathbf{z}}\mathbf{z}\mathbf{w}^\mathsf{T}$ is idempotent: for any n-vector $\mathbf{v}$,

$$\begin{aligned}
PP\mathbf{v} &= P\left(\frac{\mathbf{w}^\mathsf{T}\mathbf{v}}{\mathbf{w}^\mathsf{T}\mathbf{z}}\,\mathbf{z}\right) && \text{(by formula for } P\mathbf{v}\text{)} \\
&= \frac{\mathbf{w}^\mathsf{T}\mathbf{v}}{\mathbf{w}^\mathsf{T}\mathbf{z}}\,P\mathbf{z} && \text{(by linearity)} \\
&= \frac{\mathbf{w}^\mathsf{T}\mathbf{v}}{\mathbf{w}^\mathsf{T}\mathbf{z}}\left(\frac{\mathbf{w}^\mathsf{T}\mathbf{z}}{\mathbf{w}^\mathsf{T}\mathbf{z}}\,\mathbf{z}\right) && \text{(by formula for } P\mathbf{v} \text{ with } \mathbf{v}=\mathbf{z}\text{)} \\
&= \frac{\mathbf{w}^\mathsf{T}\mathbf{v}}{\mathbf{w}^\mathsf{T}\mathbf{z}}\,\mathbf{z}.
\end{aligned}$$

If P is a projection operator, then so is the operator Q given by

$$Q \;=\; I - P,$$

which sends $\mathbf{v}$ to $\mathbf{v} - P\mathbf{v}$. Pairs of projection operators such as P and Q, where $P + Q = I$, are said to be *complementary projectors*. Every n-vector $\mathbf{v}$ is (additively) decomposed into two "parts", $P\mathbf{v}$ and $Q\mathbf{v}$, such that

$$P\mathbf{v} + Q\mathbf{v} \;=\; \mathbf{v}.$$

Notice that $PQ\mathbf{v} = QP\mathbf{v} = \mathbf{0}$ as well.

In the case where $P = (1/(\mathbf{w}^\mathsf{T}\mathbf{z}))\mathbf{z}\mathbf{w}^\mathsf{T}$ is an elementary projector, $P\mathbf{v}$ is the "part" of $\mathbf{v}$ on the line $\{c\mathbf{z} : c \in \mathbb{R}\}$, while $Q\mathbf{v}$ is the "part" of $\mathbf{v}$ in the hyperplane $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\mathsf{T}\mathbf{x} = 0\}$. This last part is a consequence of the following theorem, since $\mathsf{NS}(P) = \mathsf{NS}(\mathbf{w}^\mathsf{T}) = \mathbb{H}$.

Theorem 1. *Let P be a projection operator on $\mathbb{R}^n$. Then $\mathsf{CS}(I-P) = \mathsf{NS}(P)$.*

Proof. We first show that $\mathsf{CS}(I-P) \subseteq \mathsf{NS}(P)$. Take any $\mathbf{v} \in \mathsf{CS}(I-P)$, so there exists an n-vector $\mathbf{x}$ such that $\mathbf{v} = (I-P)\mathbf{x}$. Since $P(I-P)\mathbf{x} = \mathbf{0}$ by idempotency of P, it follows that $P\mathbf{v} = P(I-P)\mathbf{x} = \mathbf{0}$ as well, which implies $\mathbf{v} \in \mathsf{NS}(P)$. We conclude that $\mathsf{CS}(I-P) \subseteq \mathsf{NS}(P)$.

We now show that $\mathsf{NS}(P) \subseteq \mathsf{CS}(I-P)$. Take any $\mathbf{v} \in \mathsf{NS}(P)$. This implies that $\mathbf{v} = P\mathbf{v} + (I-P)\mathbf{v} = \mathbf{0} + (I-P)\mathbf{v}$, so $\mathbf{v} \in \mathsf{CS}(I-P)$ as well. We conclude that $\mathsf{NS}(P) \subseteq \mathsf{CS}(I-P)$. $\square$

Continuing from the previous example. The projector complementary to P is $Q = I - P$. For $\mathbf{v} = (0, 2)$, we have $P\mathbf{v} = -2\mathbf{z} = (-2, 0)$ and $Q\mathbf{v} = (0, 2) - (-2, 0) = (2, 2)$. We have $P\mathbf{v} + Q\mathbf{v} = (-2, 0) + (2, 2) = (0, 2) = \mathbf{v}$. See Figure 1. ∎

Caution. Later, we will discuss a special class of projectors, "orthoprojectors", but only after introducing the prerequisite concept of "orthogonality".

1.4 Reflections

For every projection operator P, there is an associated (invertible) linear operator $W = I - 2P$ called a *reflection operator* (a.k.a. *reflector*). The operator W sends $\mathbf{v}$ to $\mathbf{v} - 2P\mathbf{v}$. Equivalently, if $Q = I - P$ is the projector complementary to P, then $W = Q - P$, so W sends $\mathbf{v}$ to $Q\mathbf{v} - P\mathbf{v}$. The key property of a reflector is $WW = I$, so W is invertible with $W^{-1} = W$. Applying W twice to a vector always gives back the same vector. Let us verify this property: for any n-vector $\mathbf{v}$,

$$\begin{aligned}
WW\mathbf{v} &= W(\mathbf{v} - 2P\mathbf{v}) && \text{(by definition of } W\text{)} \\
&= (\mathbf{v} - 2P\mathbf{v}) - 2P(\mathbf{v} - 2P\mathbf{v}) \\
&= \mathbf{v} - 2P\mathbf{v} - 2P\mathbf{v} + 4PP\mathbf{v} && \text{(by linearity of } P\text{)} \\
&= \mathbf{v} - 2P\mathbf{v} - 2P\mathbf{v} + 4P\mathbf{v} && \text{(by idempotency of } P\text{)} \\
&= \mathbf{v}.
\end{aligned}$$

An *elementary reflection operator* (a.k.a. *elementary reflector*) is a reflector $W = I - 2P$ associated with an elementary projector $P = (1/(\mathbf{w}^\intercal\mathbf{z}))\mathbf{z}\mathbf{w}^\intercal$. Recall that P is associated with a line $\{c\mathbf{z} : c \in \mathbb{R}\}$ and a hyperplane $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\intercal\mathbf{x} = 0\}$. The associated reflector sends inputs from one side of the hyperplane to the other side by subtracting a multiple of $\mathbf{z}$.

- The "part" of $\mathbf{v}$ in the hyperplane is $Q\mathbf{v}$; the "part" of $\mathbf{v}$ on the line is $P\mathbf{v}$, which is a multiple of $\mathbf{z}$.
- If $\mathbf{v}$ is on one "side" of the hyperplane, then subtracting $P\mathbf{v}$ from $Q\mathbf{v}$ gives a vector on the opposite side of the hyperplane.

  (If we had added $P\mathbf{v}$ to $Q\mathbf{v}$, we would simply get back $\mathbf{v}$.)
- If $\mathbf{v}$ is already in the hyperplane, then W does not change it at all.

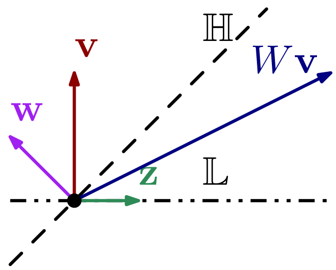

Figure 2: Applying an elementary reflector $W = I - \frac{2}{\mathbf{w}^\mathsf{T}\mathbf{z}}\mathbf{z}\mathbf{w}^\mathsf{T}$ to a vector $\mathbf{v}$.

Continuing from the previous example. For $\mathbf{v} = (0, 2)$, we have $P\mathbf{v} = (-2, 0)$, $Q\mathbf{v} = (2, 2)$, and $W\mathbf{v} = (0, 2) - 2(-2, 0) = (4, 2)$. Let $\mathbf{u} = (4, 2)$. Then $P\mathbf{u} = (2, 0)$, $Q\mathbf{u} = (2, 2)$, and $W\mathbf{u} = (4, 2) - 2(2, 0) = (0, 2) = \mathbf{v}$. See Figure 2. ▮

1.5 Dilations

For every projection operator P, there is an associated family of (invertible) linear operators $M_c = I + (c-1)P$, for every non-zero scalar c, each called a *dilation operator* (a.k.a. *dilator*). The operator M_c sends $\mathbf{v}$ to $\mathbf{v} + (c-1)P\mathbf{v}$. The case $c = -1$ gives the reflector associated with P, and the case $c = 1$ gives the identity operator. It can be verified that the inverse of M_c is $M_c^{-1} = M_{1/c}$.

An *elementary dilation operator* (a.k.a. *elementary dilator*) is a dilator $M_c = I+(c-1)P$ associated with an elementary projector $P = (1/(\mathbf{w}^\mathsf{T}\mathbf{z}))\mathbf{z}\mathbf{w}^\mathsf{T}$. Recall that if Q is the projector complementary to P, then $M_c = cP + Q$. In other words, $M_c\mathbf{v}$ scales the part of $\mathbf{v}$ that is in the line $\{c\mathbf{z} : c \in \mathbb{R}\}$ by a factor of c, and leaves alone the part of $\mathbf{v}$ that is in the hyperplane $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\mathsf{T}\mathbf{x} = 0\}$.

Continuing from the previous example. For $\mathbf{v} = (0, 2)$, we have $P\mathbf{v} = (-2, 0)$, $Q\mathbf{v} = (2, 2)$, and $M_3\mathbf{v} = 3(-2, 0) + (2, 2) = (-4, 2)$. Let $\mathbf{u} = (-4, 2)$. Then $P\mathbf{u} = (-6, 0)$, $Q\mathbf{u} = (2, 2)$, and $M_{\frac{1}{3}}\mathbf{u} = \frac{1}{3}(-6, 0) + (2, 2) = (0, 2)$. See Figure 3. ▮

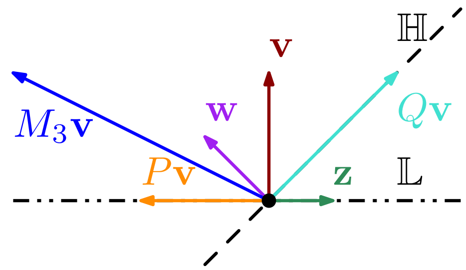

Figure 3: Applying an elementary dilator $M_3 = I + \frac{2}{\mathbf{w}^\top \mathbf{z}} \mathbf{z}\mathbf{w}^\top$ to a vector $\mathbf{v}$.

1.6 Shears

For any non-zero linear functional $\mathbf{w}^\top$ on $\mathbb{R}^n$, and any non-zero vector $\mathbf{y} \in \mathsf{NS}(\mathbf{w}^\top) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\top \mathbf{x} = 0\}$, consider the linear operator S defined by

$$S = I + \mathbf{y}\mathbf{w}^\top.$$

Such an operator is called a *shear operator* (a.k.a. *transvection operator*). Let us determine its effect on $\mathbf{v} \in \mathbb{R}^n$.

- The operator S on $\mathbf{v}$ adds a multiple of $\mathbf{y}$ to $\mathbf{v}$; the amount is proportional to how "far" the affine hyperplane $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\top \mathbf{x} = \mathbf{w}^\top \mathbf{v}\}$ (which contains $\mathbf{v}$) is from the homogeneous hyperplane $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\top \mathbf{x} = 0\}$.

- Since $\mathbf{y} \in \mathsf{NS}(\mathbf{w}^\top)$, the result $S\mathbf{v}$ remains in the affine hyperplane containing $\mathbf{v}$:

$$\mathbf{w}^\top (S\mathbf{v}) = \mathbf{w}^\top (\mathbf{v} + (\mathbf{w}^\top \mathbf{v})\mathbf{y}) = \mathbf{w}^\top \mathbf{v} + (\mathbf{w}^\top \mathbf{v})\cancelto{0}{(\mathbf{w}^\top \mathbf{y})} = \mathbf{w}^\top \mathbf{v}.$$

- For a visualization of the effect of S, imagine a stack of playing cards on a table, where the surface of the table represents $\mathsf{NS}(\mathbf{w}^\top)$, and the cards are (subsets of) other affine hyperplanes $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\top \mathbf{x} = b\}$.

  Since $\mathbf{y} \in \mathsf{NS}(\mathbf{w}^\top)$, it represents a direction contained in the plane of the table. The effect of S is to slide all of the cards in direction $\mathbf{y}$ by an amount directly proportional to the "height" of the card. So the bottom card doesn't move, and the top card moves the most.

The shear operator S is invertible, and its inverse is given by $S^{-1} = I + (-\mathbf{y})\mathbf{w}^\top$, which is also a shear operator since $-\mathbf{y} \in \mathsf{NS}(\mathbf{w}^\top)$.

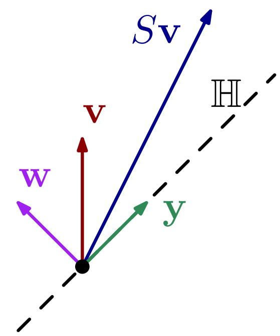

Figure 4: Applying a shear operator $S = I + \mathbf{y}\mathbf{w}^\top$ to a vector $\mathbf{v}$.

Example. Let $\mathbf{w} = (-1, 1)$ and $\mathbf{y} = (1, 1)$, and note that $\mathbf{w}^\top \mathbf{y} = 0$. Consider the shear operator $S = I + \mathbf{y}\mathbf{w}^\top$. For $\mathbf{v} = (0, 2)$, we have $S\mathbf{v} = (0, 2) + 2(1, 1) = (2, 4)$. Observe that $\mathbf{w}^\top \mathbf{v} = \mathbf{w}^\top(S\mathbf{v}) = 2$. See Figure 4. ∎

1.7 Elementary row operations redux

What do elementary reflectors, elementary dialators, and shear operators have in common?

- They are obtained by adding a matrix of rank 1 to the identity matrix.

  (Why does $\mathbf{f}\mathbf{g}^\top$ for non-zero n-vectors $\mathbf{f}$ and $\mathbf{g}$ have rank equal to 1? Answers: Each column (resp. row) of $\mathbf{f}\mathbf{g}^\top$ is a multiple of $\mathbf{f}$ (resp. $\mathbf{g}^\top$).)

- They are invertible linear operators.

- Their inverses are of the same "type".

Every elementary row operation performed by Elimination is either an elementary reflector, an elementary dialator, or a shear operator.

- To swap rows i and j, apply the elementary reflector with $\mathbf{w} = \mathbf{e}_i - \mathbf{e}_j$ and $\mathbf{z} = \mathbf{e}_j - \mathbf{e}_i$.

- To multiply row i by $c \neq 0$, apply the elementary dilator with $\mathbf{w} = \mathbf{z} = \mathbf{e}_i$ and the same scalar c.

- To subtract c times row i from row j (for $i \neq j$), apply the shear operator with $\mathbf{w} = \mathbf{e}_i$ and $\mathbf{y} = -c\mathbf{e}_j$.

Combining this with the fact that every invertible matrix is a product of elementary matrices, we have the following theorem.

Theorem 2. *Every invertible matrix is the product of some sequence of elementary reflectors, elementary dilators, and shear operators.*

2 Representations of linear transformations

You are likely familiar with many other linear transformations, and perhaps even applied them to some of your favorite vectors.

Example. Consider the vector space $\mathsf{P}_d(\mathbb{R})$ of polynomials with real coefficients of degree at most d. The *differentiation operator* $\frac{\mathrm{d}}{\mathrm{d}t} \colon \mathsf{P}_d(\mathbb{R}) \to \mathsf{P}_d(\mathbb{R})$ sends a polynomial to its derivative:

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigl(a_0\, t^0 + a_1\, t + a_2\, t^2 + \cdots + a_d\, t^d\bigr) \;=\; a_1 + 2a_2\, t + \cdots + da_d\, t^{d-1}.$$

It is linear (as one may remember from calculus), since

$$\frac{\mathrm{d}}{\mathrm{d}t}(c\, f(t) + g(t)) \;=\; c\,\frac{\mathrm{d}}{\mathrm{d}t}(f(t)) + \frac{\mathrm{d}}{\mathrm{d}t}(g(t))$$

for any scalar c and polynomials $f(t), g(t) \in \mathsf{P}_d(\mathbb{R})$. ∎

As we mentioned before, it can be helpful to work with coordinates with respect to an ordered basis when dealing with general finite-dimensional vector spaces (such as spaces of polynomials or functions). And we previously saw how to change between coordinate systems given by different ordered bases. But how do we apply linear transformations to these vectors?

If $T \colon \mathbb{V} \to \mathbb{W}$ is a linear transformation between finite-dimensional vector spaces $\mathbb{V}$ and $\mathbb{W}$ with ordered bases $\mathcal{F}$ (for $\mathbb{V}$) and $\mathcal{H}$ (for $\mathbb{W}$), then there is a matrix $[T]_{\mathcal{F}\to\mathcal{H}}$ such that, for any $\mathbf{v} \in \mathbb{V}$,

$$[T(\mathbf{v})]_{\mathcal{H}} \;=\; [T]_{\mathcal{F}\to\mathcal{H}}\, [\mathbf{v}]_{\mathcal{F}}.$$

This is the *matrix representation* of T with respect to input space basis $\mathcal{F}$ and output space basis $\mathcal{H}$.[1] We previously established this fact for the cases

[1]Sometimes the notation $[T]_{\mathcal{F}}^{\mathcal{H}}$ is used instead of $[T]_{\mathcal{F}\to\mathcal{H}}$.

where $\mathbb{V}$ and $\mathbb{W}$ are Cartesian spaces and the ordered bases are the standard ordered bases. Here is the matrix for the general case: letting $\mathcal{F} = (\mathbf{f}_1, \dots, \mathbf{f}_n)$,

$$[T]_{\mathcal{F}\to\mathcal{H}} = \begin{bmatrix} [T(\mathbf{f}_1)]_{\mathcal{H}} & \cdots & [T(\mathbf{f}_n)]_{\mathcal{H}} \end{bmatrix}.$$

So, the jth column is obtained by applying the transformation T to jth vector in the <u>*input space basis*</u> $\mathcal{F}$, and then applying the standard coordinate map for the <u>*output space basis*</u> $\mathcal{H}$.

Continuing the previous example. An ordered basis for $\mathsf{P}_3(\mathbb{R})$ is $\mathcal{F} = (1, t, t^2, t^3)$. The matrix representing $\frac{\mathrm{d}}{\mathrm{d}t} \colon \mathsf{P}_3(\mathbb{R}) \to \mathsf{P}_3(\mathbb{R})$, where the input space basis and output space basis are both $\mathcal{F}$, is

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}\right]_{\mathcal{F}\to\mathcal{F}} = \begin{bmatrix} \left[\frac{\mathrm{d}}{\mathrm{d}t}(1)\right]_{\mathcal{F}} & \left[\frac{\mathrm{d}}{\mathrm{d}t}(t)\right]_{\mathcal{F}} & \left[\frac{\mathrm{d}}{\mathrm{d}t}(t^2)\right]_{\mathcal{F}} & \left[\frac{\mathrm{d}}{\mathrm{d}t}(t^3)\right]_{\mathcal{F}} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

If the output space basis is changed to $\mathcal{H} = (1, t, t^2 - 1, t^3 - 3t)$, then

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}\right]_{\mathcal{F}\to\mathcal{H}} = \begin{bmatrix} \left[\frac{\mathrm{d}}{\mathrm{d}t}(1)\right]_{\mathcal{H}} & \left[\frac{\mathrm{d}}{\mathrm{d}t}(t)\right]_{\mathcal{H}} & \left[\frac{\mathrm{d}}{\mathrm{d}t}(t^2)\right]_{\mathcal{H}} & \left[\frac{\mathrm{d}}{\mathrm{d}t}(t^3)\right]_{\mathcal{H}} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 3 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that we could have obtained this by multiplying $[\frac{\mathrm{d}}{\mathrm{d}t}]_{\mathcal{F}\to\mathcal{F}}$ on the left by the change-of-coordinates matrix $[\mathrm{id}]_{\mathcal{F}\to\mathcal{H}}$:

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}\right]_{\mathcal{F}\to\mathcal{H}} = [\mathrm{id}]_{\mathcal{F}\to\mathcal{H}} \left[\frac{\mathrm{d}}{\mathrm{d}t}\right]_{\mathcal{F}\to\mathcal{F}} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here, id denotes the identity operator (in this context, for $\mathsf{P}_3(\mathbb{R})$). ∎

Both the input and output bases are important to mind when representing a linear transformation as a matrix. In fact, as we'll see later in the course, sometimes judicious choices of bases can result in a very simple matrix.

Example. Consider the linear operator $T \colon \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{bmatrix} ,$$

so

$$[T]_{\mathcal{E}_3 \to \mathcal{E}_3} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} ,$$

where $\mathcal{E}_3$ is the standard ordered basis for $\mathbb{R}^3$. Consider another ordered basis $\mathcal{F} = ((1, -1, 0), (1, 1, -2), (1, 1, 1))$. We have

$$[\mathrm{id}]_{\mathcal{F} \to \mathcal{E}_3} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -2 & 1 \end{bmatrix} , \quad [\mathrm{id}]_{\mathcal{E}_3 \to \mathcal{F}} = \begin{bmatrix} 1/2 & -1/2 & 0 \\ 1/6 & 1/6 & -1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} .$$

Then

$$[T]_{\mathcal{F} \to \mathcal{F}} = [\mathrm{id}]_{\mathcal{E}_3 \to \mathcal{F}} \, [T]_{\mathcal{E}_3 \to \mathcal{E}_3} \, [\mathrm{id}]_{\mathcal{F} \to \mathcal{E}_3} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix} ,$$

a diagonal matrix. So, in terms of the "coordinates" with respect to the basis $\mathcal{F}$, the linear transformation T just scales each coordinate of the input separately. ∎

A Form of elementary projection operators

Given non-zero vectors $\mathbf{w}, \mathbf{z} \in \mathbb{R}^n$ such that $\mathbf{w}^\top \mathbf{z} \neq 0$, we want to find a linear operator P on $\mathbb{R}^n$ such that

1. $P^2 = P$ (i.e., P is idempotent),
2. $\mathsf{CS}(P) = \operatorname{span}(\{\mathbf{z}\})$, and
3. $\mathsf{NS}(P) = \mathsf{NS}(\mathbf{w}^\top)$.

The second property tells us that every column of P should be a multiple of $\mathbf{z}$. The third property tells us that the homogeneous system of linear equations $P\mathbf{x} = \mathbf{0}$ should have the same solution set as $\mathbf{w}^\top \mathbf{x} = 0$. So P and $\mathbf{w}^\top$ should have the same R matrix in their CR factorizations. Since $\mathbf{w} \neq \mathbf{0}$, the CR factorization of $\mathbf{w}^\top$ has

$$C = \begin{bmatrix} d \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} \leftarrow & \frac{1}{d}\mathbf{w}^\top & \rightarrow \end{bmatrix},$$

where d is the first non-zero component of $\mathbf{w} = (w_1, \dots, w_n)$. In addition, since $\mathbf{z} \neq \mathbf{0}$, the CR factorization of P has

$$C = \begin{bmatrix} a\,\mathbf{z} \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} \leftarrow & \frac{1}{d}\mathbf{w}^\top & \rightarrow \end{bmatrix},$$

where $a\mathbf{z}$ is the first non-zero column of P, and d is as given above. So

$$P = (a\,\mathbf{z})\left(\frac{1}{d}\mathbf{w}^\top\right) = \frac{a}{d}\mathbf{z}\mathbf{w}^\top.$$

In other words, P is a scalar multiple $c = a/d$ of $\mathbf{z}\mathbf{w}^\top$. It remains to determine this scalar c. Since $\mathsf{CS}(P) = \operatorname{span}(\{\mathbf{z}\})$ by the second property, there exists $\mathbf{x} \in \mathbb{R}^n$ such that

$$\mathbf{z} = P\mathbf{x}.$$

Applying P to both sides gives

$$P\mathbf{z} = P^2\mathbf{x}.$$

But by the first property, $P^2\mathbf{x} = P\mathbf{x}$, which equals $\mathbf{z}$. Hence $\mathbf{z} = P\mathbf{z}$. So, using the form $P = c\mathbf{z}\mathbf{w}^\top$, we have

$$\mathbf{z} = P\mathbf{z} = (c\,\mathbf{z}\mathbf{w}^\top)\mathbf{z} = (c\,\mathbf{w}^\top\mathbf{z})\,\mathbf{z}.$$

Since $\mathbf{z} \neq \mathbf{0}$, it follows by the Unique Representations Theorem that

$$c\,\mathbf{w}^{\mathsf{T}}\mathbf{z} \;=\; 1,$$

i.e., $c = 1/\mathbf{w}^{\mathsf{T}}\mathbf{z}$. We conclude

$$P \;=\; \frac{1}{\mathbf{w}^{\mathsf{T}}\mathbf{z}}\,\mathbf{z}\mathbf{w}^{\mathsf{T}}.$$