$\mathbb{R}^n$, LINEAR TRANSFORMATIONS AND MATRICES
BASED ON LECTURE NOTES BY JAMES MCKERNAN

Blackboard 1. *A vector in* $\mathbb{R}^n$ *is an* n*-tuple* $\vec{v} = (v_1, v_2, \dots, v_n)$. *Given* $\vec{v}, \vec{w} \in \mathbb{R}^n$, *their sum is*

$$\vec{v} + \vec{w} = (v_1 + w_1, v_2 + w_2, \dots, v_n + w_n),$$

their dot product is

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + \dots + v_n w_n,$$

and for $\lambda \in \mathbb{R}$

$$\lambda \vec{v} = (\lambda v_1, \lambda v_2, \dots, \lambda v_n).$$

The norm of $\vec{v}$ *is*

$$\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}.$$

Blackboard 2. *The standard basis of* $\mathbb{R}^n$ *is the set of vectors,*

$$\hat{e}_1 = (1, 0, \dots, 0), \quad \hat{e}_2 = (0, 1, \dots, 0), \quad \dots, \hat{e}_n = (0, 0, \dots, 1).$$

If $\vec{v} = (v_1, v_2, \dots, v_n)$, *then*

$$\vec{v} = v_1 \hat{e}_1 + v_2 \hat{e}_2 + \dots + v_n \hat{e}_n.$$

Let's adopt the (somewhat ad hoc) convention that $\vec{v}$ and $\vec{w}$ are parallel if and only if either $\vec{v}$ is a scalar multiple of $\vec{w}$, or vice-versa. Note that if both $\vec{v}$ and $\vec{w}$ are non-zero vectors, then $\vec{v}$ is a scalar multiple of $\vec{w}$ if and only if $\vec{w}$ is a scalar multiple of $\vec{v}$.

Theorem 3 (Cauchy-Schwarz-Bunyakowski)**.** *If* $\vec{v}, \vec{w} \in \mathbb{R}^n$ *then*

$$|\vec{v} \cdot \vec{w}| \leq \|v\| \|w\|,$$

with equality iff $\vec{v}$ *is parallel to* $\vec{w}$.

Proof. If either $\vec{v}$ or $\vec{w}$ is the zero vector, then there is nothing to prove. So we may assume that neither vector is the zero vector.

Let $\vec{u} = x\vec{v} + \vec{w}$, where x is a scalar, and let

$$f(x) = \vec{u} \cdot \vec{u} = |\vec{u}|^2.$$

Then

$$0 \leq f(x) = (\vec{v} \cdot \vec{v})x^2 + 2(\vec{v} \cdot \vec{w})x + \vec{w} \cdot \vec{w} = ax^2 + bx + c.$$

So $f(x)$ has at most one root. It follows that the discriminant $b^2 - 4ac \leq 0$, with equality iff $f(x)$ has a root. Hence

$$b^2 - 4ac = 4(\vec{v} \cdot \vec{w})^2 - 4\|\vec{v}\|^2 \|\vec{w}\|^2 \leq 0.$$

Rearranging, gives

$$(\vec{v} \cdot \vec{w})^2 \leq \|\vec{v}\|^2 \|\vec{w}\|^2.$$

Taking square roots, gives

$$|\vec{v} \cdot \vec{w}| \leq \|v\| \|w\|.$$

Now, $f(x)$ has a root λ iff we have equality here. Hence $0 = f(\lambda) = |\lambda \vec{v} + \vec{w}|^2$, and $\lambda \vec{v} + \vec{w} = \vec{0}$. In other words, $\vec{w} = -\lambda \vec{v}$ and $\vec{v}$ and $\vec{w}$ are parallel. $\square$

Blackboard 4. *If $\vec{v}$ and $\vec{w} \in \mathbb{R}^n$ are non-zero vectors, then the angle between them is the unique angle $0 \leq \theta \leq \pi$ such that*

$$\cos\theta = \frac{\vec{v} \cdot \vec{w}}{\|\vec{v}\|\|\vec{w}\|}.$$

Note that the fraction is between -1 and 1, by the Cauchy-Schwarz-Bunjakowski inequality, so this does makes sense. We also showed in that the angle is 0 or π if and only if $\vec{v}$ and $\vec{w}$ are parallel.

Blackboard 5. *A linear transformation $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a function satisfying*

$$f(\lambda\vec{v}) = \lambda f(\vec{v})$$

and

$$f(\vec{v} + \vec{w}) = f(\vec{v}) + f(\vec{w}).$$

It doesn't matter if we apply the function before or after multiplying by a scalar. It also doesn't matter if we apply the function before or after adding.

Theorem 6. *A linear transformation $f\colon \mathbb{R} \to \mathbb{R}$ is of the form $f(x) = ax$ where $a = f(1)$.*

Proof.

$$f(x) = f(x \cdot 1) = x \cdot f(1) = ax.$$

$\square$

Theorem 7. *A linear transformation $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is of the form*

$$f(\vec{v}) = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 \\ a_{21}v_1 + a_{22}v_2 \end{pmatrix},$$

where

$$\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} = f(\hat{e}_1)$$

and

$$\begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} = f(\hat{e}_2),$$

or, equivalently, $a_{ij} = f(\hat{e}_j)_i = \hat{e}_i \cdot f(\hat{e}_j)$.

Proof. By the definition of a linear transformation,

$$\begin{aligned} f(\vec{v}) &= f(v_1\hat{e}_1 + v_2\hat{e}_2) \\ &= f(v_1\hat{e}_1) + f(v_2\hat{e}_2) \\ &= v_1 f(\hat{e}_1) + v_2 f(\hat{e}_2). \end{aligned}$$

substituting the definitions of the a_{ij}'s, we get

$$\begin{aligned} &= v_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + v_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}v_1 + a_{12}v_2 \\ a_{21}v_1 + a_{22}v_2 \end{pmatrix}. \end{aligned}$$

$\square$

Theorem 8. *A linear transformation $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is of the form*

$$f(\vec{v}) = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + \ldots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \ldots + a_{2n}v_n \\ \ldots \\ a_{m1}v_1 + a_{m2}v_2 + \ldots + a_{mn}v_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1j}v_j \\ \sum_{j=1}^n a_{2j}v_j \\ \ldots \\ \sum_{j=1}^n a_{mj}v_j \end{pmatrix},$$

where $a_{ij} = f(\hat{e}_j)_i = \hat{e}_i \cdot f(\hat{e}_j)$.

The proof is the same as for the previous case. One way to describe such a transformation is in a matrix.

Blackboard 9. *The $m \times n$ matrix A associated with the transformation $f\colon \mathbb{R}^n \to \mathbb{R}^m$ the m-by-a array of real numbers*

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

where $a_{ij} = f(\hat{e}_j)_i = \hat{e}_i \cdot f(\hat{e}_j)$. We denote $A = (a_{ij})$.

Example:

Blackboard 10. *If $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is*

$$f(\vec{v}) = \begin{pmatrix} 2v_1 - v_2 \\ v_1 \end{pmatrix}$$

then the associated matrix is

$$A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}.$$

Blackboard 11. *Let $A = (a_{ij})$ be the matrix associated with the linear transformation $f\colon \mathbb{R}^n \to \mathbb{R}^m$. We define the product of A with $\vec{v} \in \mathbb{R}^n$ by*

$$A \cdot \vec{v} = f(\vec{v}) = \begin{pmatrix} \sum_{j=1}^n a_{1j}v_j \\ \sum_{j=1}^n a_{2j}v_j \\ \ldots \\ \sum_{j=1}^n a_{mj}v_j \end{pmatrix}.$$

The is the same as saying that the ith component of $A \cdot \vec{v}$ is the "dot product" of the ith row of A with $\vec{v}$.

Theorem 12. *Let $f, g\colon \mathbb{R}^n \to \mathbb{R}^m$ and $h\colon \mathbb{R}^m \to \mathbb{R}^p$ be linear transformations. Then*

- *$f_1(\vec{v}) = \lambda f(\vec{v})$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$.*
- *$f_2(\vec{v}) = f(\vec{v}) + g(\vec{v})$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$.*
- *$f_3(\vec{v}) = h(f(\vec{v}))$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^p$. We also denote it by $h \circ f$.*

Theorem 13. *Let $f, g\colon \mathbb{R}^n \to \mathbb{R}^m$ and $h\colon \mathbb{R}^m \to \mathbb{R}^p$ be linear transformations. Let $A = (a_{ij})$, $B = (b_{ij})$ and $C = (c_{ij})$ be the matrices associated with f, g and h, respectively.*

- *Let $f_1(\vec{v}) = \lambda f(\vec{v})$. Then the matrix associated with f_1 is (λa_{ij}). (We denote this matrix by λA.)*

- *$f_2(\vec{v}) = f(\vec{v}) + g(\vec{v})$. Then the matrix associated with f_2 is $(a_{ij} + b_{ij})$. (We denote this matrix by $A + B$.)*
- *$f_3 = h(f(\vec{v}))$. Then the matrix associated with f_3 is (d_{ij}), where*
$$d_{ij} = \sum_{k=1}^{m} c_{ik} a_{kj}.$$
(We denote this matrix by $C \cdot A$.)

Multiplying a matrix A by λ means multiplying each entry by λ. Any matrix can be multiplied by any scalar.

Adding two matrices means adding the corresponding entires. Only matrices of the same dimensions can be added.

The ijth entry of $C \cdot A$ is the "dot product" of the i row of C with the jth column of A. The product CA exists only if the number of columns in C is equal to the number of rows in A.

Proof. We prove (3). Let $D = (d_{ij})$ be the matrix associated with f_3. Then by definition

$$\begin{aligned} d_{ij} &= \hat{e}_i \cdot f_3(\hat{e}_j) \\ &= \hat{e}_i \cdot h(f(\hat{e}_j)) \\ &= \hat{e}_i \cdot h(a_{1j}\hat{e}_1 + a_{2j}\hat{e}_2 + \ldots + a_{mj}\hat{e}_m) \\ &= \hat{e}_i \cdot \sum_{k=1}^{m} a_{kj} h(\hat{e}_k) \\ &= \sum_{k=1}^{m} c_{ik} a_{kj}. \end{aligned}$$

□

Example.

Blackboard 14. *If*

$$A = \begin{pmatrix} 1 & -1 \\ 3 & -4 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix},$$

then

$$A + B = \begin{pmatrix} 2 & 0 \\ 5 & -5 \end{pmatrix},$$

and

$$3A = \begin{pmatrix} 3 & -3 \\ 9 & -12 \end{pmatrix}.$$

Another example.

Blackboard 15. *Let*

$$c = \begin{pmatrix} 1 & -2 & 1 \\ 1 & -1 & 5 \end{pmatrix} \qquad \text{and} \qquad A = \begin{pmatrix} 2 & 1 \\ 1 & -4 \\ -1 & 1 \end{pmatrix}.$$

Then $D = CA$ has shape 2×2, and in fact

$$D = CA = \begin{pmatrix} -1 & 10 \\ -4 & 10 \end{pmatrix}.$$

Theorem 16. *Let A, B and C be matrices with dimensions $m \times n$, $n \times p$ and $p \times q$, respectively. Then $(AB)C = A(BC)$.*

Matrix multiplication is associative.

Proof. Let f, g, h be the linear transformations associated with A, B and C, respectively. Then AB is the matrix associated with $f \circ g$, and BC is the matrix associated with $g \circ h$. Hence $(AB)C$ is the matrix associated with $(f \circ g) \circ h$, and $A(BC)$ is the matrix associated with $f \circ (g \circ h)$. But $(f \circ g) \circ h = f \circ (g \circ h)$, and therefore $A(BC) = (AB)C$. $\square$

Blackboard 17. *In general, $AB \neq BA$. For example, if $A = (1, 2)$ and $B = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ then*

$$AB = (1, 2) \cdot \begin{pmatrix} 3 \\ 4 \end{pmatrix} = (11) \quad \textit{and} \quad BA = \begin{pmatrix} 3 \\ 4 \end{pmatrix} \cdot (1, 2) = \begin{pmatrix} 3 & 6 \\ 4 & 8 \end{pmatrix}.$$

Even if A and B are the same size then sometimes $AB \neq BA$. Suppose

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \textit{and} \qquad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Then AB and BA are both 2×2 matrices. But

$$AB = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \qquad \textit{and} \qquad BA = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

This is related to the fact that in general $f \circ g \neq g \circ f$.