Math 51
Lecture 13 – July 18, 2023

Topic(s): linear functions, matrices

A scalar-valued function $f : \mathbb{R}^n \to \mathbb{R}$ is called

- **affine** if it has the form $f(x_1, \ldots, x_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n + b$ for some scalars $a_1, \ldots, a_n, b$. Note that $f(\mathbf{0}) = b$.
- **linear** if it has the form $f(x_1, \ldots, x_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ for some scalars $a_1, \ldots, a_n$. Note that a linear function is an affine function with the added condition $f(\mathbf{0}) = 0$.

Similarly, a vector-valued function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ ($\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$) is called

- **affine** if each of its component functions $f_i : \mathbb{R}^n \to \mathbb{R}$ is affine.
- **linear** if each of its component functions $f_i : \mathbb{R}^n \to \mathbb{R}$ is linear.

Linear $\Rightarrow$ Affine

Example 1. Consider the function $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$\mathbf{f}(x, y, z) = \begin{bmatrix} x - y + z + 3 \\ z - x \\ y + x + 1 \end{bmatrix}.$$

$f_1(x,y,z) = (1)x + (-1)y + (1)z + 3$
$f_2(x,y,z) = (-1)x + (0)y + (1)z + 0$
$f_3(x,y,z) = (1)x + (1)y + (0)z + 1$

Is $\mathbf{f}$ affine, linear, or neither?

f_1 and f_3 are affine, and f_2 is linear. So, $\vec{f}$ is affine.

How about the function $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$\mathbf{g}(x, y, z) = \begin{bmatrix} x - y + z \\ z - x \\ y + x \end{bmatrix}?$$

$g_1(x,y,z) = (1)x + (-1)y + (1)z$
$g_2(x,y,z) = (-1)x + (0)y + (1)z$
$g_3(x,y,z) = (1)x + (1)y + (0)z$

Since g_1, g_2, g_3 are all linear, $\vec{g}$ is linear.

How about the function $\mathbf{h} : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$\mathbf{h}(x, y, z) = \begin{bmatrix} x + 2y - 3z \\ xyz \end{bmatrix}?$$

$h_1(x,y,z) = (1)x + (2)y + (-3)z$
$h_2(x,y,z) = xyz$

Since h_2 is neither linear nor affine, $\vec{h}$ is neither linear nor affine.

An affine function $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^3$ can be written out as

$$\mathbf{f}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} a_{1,1}x + a_{1,2}y + a_{1,3}z + a_{1,4} \\ a_{2,1}x + a_{2,2}y + a_{2,3}z + a_{2,4} \\ a_{3,1}x + a_{3,2}y + a_{3,3}z + a_{3,4} \end{bmatrix}$$

Similarly, a linear function $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^3$ can be written out as

$$\mathbf{f}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} a_{1,1}x + a_{1,2}y + a_{1,3}z \\ a_{2,1}x + a_{2,2}y + a_{2,3}z \\ a_{3,1}x + a_{3,2}y + a_{3,3}z \end{bmatrix}$$

As we can see, there are a lot of coefficients to keep track of, and it is easier to organize these in matrices.

An $m \times n$ **matrix** is a rectangular array A of numbers presented like this:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}.$$

The collection of entries $\begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,n} \end{bmatrix}$ is called the i**th row**, and the collection of entries $\begin{bmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{bmatrix}$ is called the j**th column**. The entry is ith row and jth column, $a_{i,j}$, is called the ij**-entry** or (i, j)**-entry**.

If A is an $m \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n$, the **matrix-vector product** $A\mathbf{x} \in \mathbb{R}^m$ is defined by

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

In other words, if $\mathbf{r}_1, \ldots, \mathbf{r}_m$ represent the rows of A (these are n-vectors), then

$$A\mathbf{x} = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{x} \\ \mathbf{r}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{r}_m \cdot \mathbf{x} \end{bmatrix}.$$

Proposition 13.3.8. A function $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{R}^m$ is linear when $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ for an $m \times n$ matrix A.

We can also use matrix-vector products to encode systems of linear equations.

Example 2. For each of the following functions $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, show it is linear or affine by writing it as $\mathbf{f}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ or explain why it is neither.

(a) $\mathbf{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -x - y \\ -x + 3y - 2 \end{bmatrix} = \begin{bmatrix} -x-y \\ -x+3y \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \end{bmatrix}$

$\vec{f}$ is affine.

(b) $\mathbf{f}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 3x_1x_2 \\ x_1 - x_2 \end{bmatrix}$ $f_1(x_1, x_2) = 3x_1x_2$ is neither affine nor linear.

$\vec{f}$ is neither affine nor linear.

(c) $\mathbf{f}\left(\begin{bmatrix} v \\ w \end{bmatrix}\right) = \begin{bmatrix} v - 2w \\ 3v \end{bmatrix} = \begin{bmatrix} v-2w \\ 3v \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 3 & 0 \end{bmatrix}\begin{bmatrix} v \\ w \end{bmatrix}$

$\vec{f}$ is linear.

Theorem 13.4.1. If $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ are the columns of A, i.e. $A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix}$, then

$$A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots x_n\mathbf{c}_n \in \mathbb{R}^m.$$

In particular, the matrix-vector product is a linear combination of the columns of A, where the coefficients are the entries of the x-vector.

Theorem 13.4.5. For a linear function $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, the matrix A has columns $\mathbf{f}(\mathbf{e}_1), \mathbf{f}(\mathbf{e}_2), \ldots, \mathbf{f}(\mathbf{e}_n)$, where $\mathbf{e}_i$ is the ith coordinate vector or the ith standard basis vector in $\mathbb{R}^n$. In other words, $A\mathbf{e}_j$ is the jth column of A. Hence, given $\mathbf{f}$, we can reconstruct A by using the standard basis vectors.

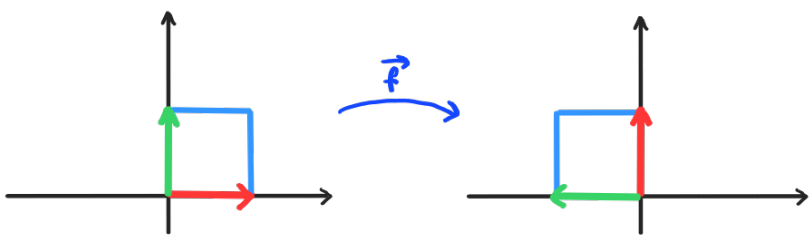
Example 3. Let us revisit the function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathbf{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -y \\ x \end{bmatrix}.$$

Recall that this takes a 2-vector and rotates it 90 degrees counterclockwise.

Since $\vec{f}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\vec{f}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$, $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Note that

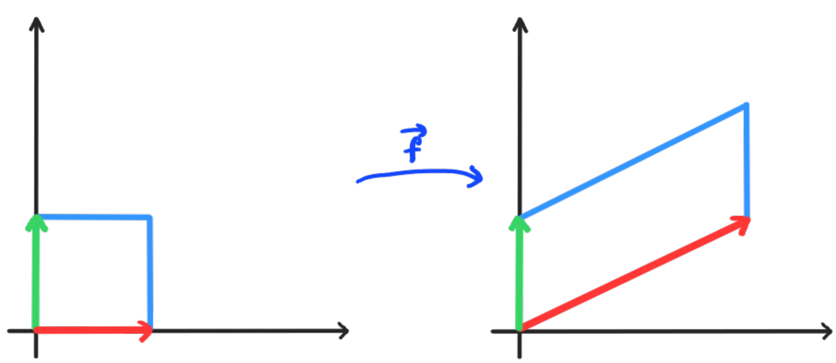
$$A\vec{x} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ -x \end{bmatrix} = \vec{f}(\vec{x}).$$



Example 4. Consider the function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathbf{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x \\ x + y \end{bmatrix}.$$

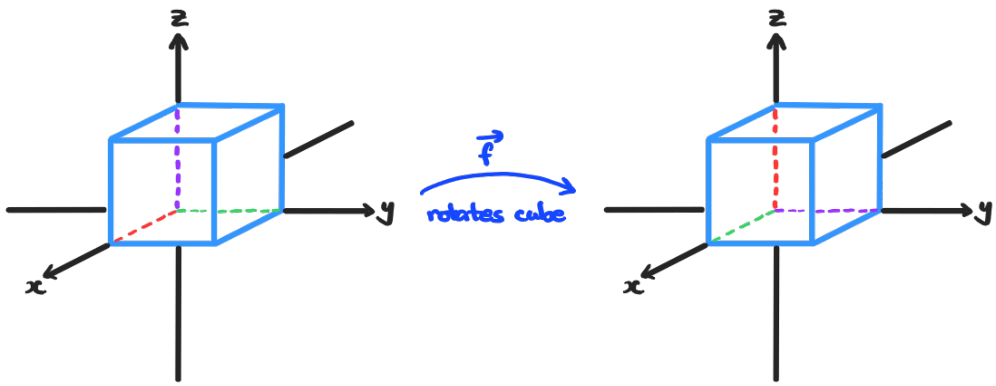
Since $\vec{f}(\vec{x}) = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}\vec{x}$, $\vec{f}(\vec{e}_1) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{f}(\vec{e}_2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.



Example 5. Consider the function $\mathbf{f}: \mathbb{R}^3 \to \mathbb{R}^3$ given by $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, where

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

$\vec{f}(\vec{e}_1) = \vec{e}_3, \quad \vec{f}(\vec{e}_2) = \vec{e}_1, \quad \vec{f}(\vec{e}_3) = \vec{e}_2.$



Let $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$ be a vector-valued function $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$ with scalar-valued component functions $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$. The **derivative matrix** of $\mathbf{f}$ at $\mathbf{a} \in \mathbb{R}^n$ is the $m \times n$ matrix

$$(D\mathbf{f})(\mathbf{a}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \frac{\partial f_m}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{bmatrix}$$

The $(D\mathbf{f})(\mathbf{a})$ is also called the **Jacobian matrix**. In general, the ith row of $(D\mathbf{f})(\mathbf{a})$ is $\nabla f_i(\mathbf{a})$ written horizontally.

Example 6. For each of the following functions $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$, compute $Df(\mathbf{x})$ as an $m \times n$ matrix whose entries are functions of $\mathbf{x} \in \mathbb{R}^n$.

(a) $\mathbf{f}([t]) = \begin{bmatrix} \sin(t) \\ t^2 + 3 \\ -t \end{bmatrix}$

$$D\vec{f}(t) = \begin{bmatrix} \cos t \\ 2t \\ -1 \end{bmatrix}$$

(b) $\mathbf{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} h(2x+y) \\ h\left(\frac{x}{y}\right) \end{bmatrix}$ for a function $h: \mathbb{R} \to \mathbb{R}$ (your answer should involve h')

$$D\vec{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2h'(2x+y) & h'(2x+y) \\ \frac{1}{y} h'\left(\frac{x}{y}\right) & -\frac{x}{y^2} h'\left(\frac{x}{y}\right) \end{bmatrix}$$

Theorem 13.5.8. The best linear approximation to $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$ at $\mathbf{a} \in \mathbb{R}^n$ is given by the $m \times n$ derivative matrix $Df(\mathbf{a})$. We have the optimal approximation of m-vectors

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{a}) + (D\mathbf{f}(\mathbf{a}))(\mathbf{x} - \mathbf{a})$$

for n-vectors $\mathbf{x}$ near $\mathbf{a}$. Equivalently,

$$\mathbf{f}(\mathbf{a} + \mathbf{h}) \approx \mathbf{f}(\mathbf{a}) + (D\mathbf{f}(\mathbf{a}))\mathbf{h}$$

for n-vectors $\mathbf{h}$ near $\mathbf{0}$.

Example 7. For $\mathbf{f}: \mathbb{R}^3 \to \mathbb{R}^2$ defined by $\mathbf{f}(x, y, z) = (x^2 - y, z^3 + xy)$, work out the linear approximations for $\mathbf{a}$ near $(1, 1, 1)$ and for $\mathbf{f}(1 + h_1, 1 + h_2, 1 + h_3)$.

We get $D\vec{f}(x,y,z) = \begin{bmatrix} 2x & -1 & 0 \\ y & x & 3z^2 \end{bmatrix}$ and $D\vec{f}(1,1,1) = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 1 & 3 \end{bmatrix}$.

For $\vec{a} = (a, b, c)$ near $(1,1,1)$,

$$\vec{f}(\vec{a}) \approx \vec{f}(1,1,1) + \begin{bmatrix} 2 & -1 & 0 \\ 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} a-1 \\ b-1 \\ c-1 \end{bmatrix}$$

$$= \begin{bmatrix} 2a - b & -1 \\ a + b + 3c - 3 \end{bmatrix}.$$

For small h_1, h_2, h_3,

$$\vec{f}(1+h_1, 1+h_2, 1+h_3) \approx \vec{f}(1,1,1) + \begin{bmatrix} 2 & -1 & 0 \\ 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

$$= \begin{bmatrix} 2h_1 - h_2 \\ h_1 + h_2 + 3h_3 + 2 \end{bmatrix}.$$

Example 8. Consider the affine functions $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ and $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$\mathbf{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y-1 \\ 3x \\ y-2 \end{bmatrix} \qquad \text{and} \qquad \mathbf{g}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x-y-z+3 \\ 3x-z+2 \end{bmatrix}.$$

(a) Evaluate $(\mathbf{f} \circ \mathbf{g})\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) \in \mathbb{R}^3$ by plugging $\mathbf{g}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) \in \mathbb{R}^2$ into $\mathbf{f}$, and write it in the usual form $A\mathbf{x} + \mathbf{b}$ for a 3×3 matrix A and vector $\mathbf{b}$ (so $\mathbf{f} \circ \mathbf{g}$ is affine).

$$(\vec{f} \circ \vec{g})(x,y,z) = \vec{f}(x-y-z+3,\ 3x-z+2)$$

$$= \begin{bmatrix} (x-y-z+3)+(3x-z+2)-1 \\ 3(x-y-z+3) \\ (3x-z+2)-2 \end{bmatrix}$$

$$= \begin{bmatrix} 4x-y-2z+4 \\ 3x-3y-3z+9 \\ 3x \quad -z \end{bmatrix}$$

$$= \begin{bmatrix} 4 & -1 & -2 \\ 3 & -3 & -3 \\ 3 & 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 4 \\ 9 \\ 0 \end{bmatrix}.$$

(b) Compute $(\mathbf{f} \circ \mathbf{g})\left(\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}\right)$ in two ways: directly computing the 2-vector $\mathbf{g}\left(\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}\right)$ and plugging that into $\mathbf{f}$, and by plugging $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ into the "$A\mathbf{x} + \mathbf{b}$" form that you determined for $\mathbf{f} \circ \mathbf{g}$ in (a). Your two answers should agree!

We get $\vec{g}(1,1,-2) = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$, and so, $(\vec{f} \circ \vec{g})(1,1,-2) = \vec{f}(5,7) = \begin{bmatrix} 11 \\ 15 \\ 5 \end{bmatrix}$.

From (a), $(\vec{f} \circ \vec{g})(1,1,-2) = \begin{bmatrix} 4 & -1 & -2 \\ 3 & -3 & -3 \\ 3 & 0 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + \begin{bmatrix} 4 \\ 9 \\ 0 \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix} + \begin{bmatrix} 4 \\ 9 \\ 0 \end{bmatrix} = \begin{bmatrix} 11 \\ 15 \\ 5 \end{bmatrix}$.