Lecture 13 - Linear Functions, Matrices, and the Derivative Matrix

October 26, 2022

Goals: Distinguish linear functions from more general functions, multiply matrices by vectors, compute the derivative matrix, and compute a local approximation from a derivative matrix.

We will now dive deeper into specific kinds of functions and their properties. We start with:

<u>**Definition:**</u> A scalar-valued function $f : \mathbb{R}^n \to \mathbb{R}$ is called

- **affine** if it has the form $f(x_1, \ldots, x_n) = a_1x_1 + \cdots + a_nx_n + b$ for scalars $a_1, \ldots, a_n, b$ (in particular, $b = f(\mathbf{0})$).
- **linear** if it has the form $f(x_1, \ldots, x_n) = a_1x_1 + \cdots + a_nx_n$ for scalars $a_1, \ldots, a_n$, i.e. it is affine with $b = 0$.

requires that highest power is 1, can't have x^2

linear contains origin

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

if it has a mix of affine & linear component, it is affine

every linear function is affine

Example 1: Are the following functions affine, linear, or neither?

$$\text{(a) } f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x - y \\ \pi x + 6y \\ 3y \end{bmatrix} \qquad \text{(b) } f(x) = 2x + 1 \qquad \text{(c) } f(x) = \begin{bmatrix} x + 1 \\ 3x - 2 \\ x^2 + x \end{bmatrix}$$

a) $f: \mathbb{R}^2 \to \mathbb{R}^3$

$f_1 = 2x - y \to$ linear

$f_2 = \pi x + 6y \to$ linear

$f_3 = 3y \to$ linear

f is linear

b) affine

$y = mx$ is linear

$y = mx + b$ is affine

b) $f: \mathbb{R} \to \mathbb{R}^3$

$f_1 = x + 1 \to$ affine

$f_2 = 3x - 2 \to$ affine

$f_3 = x^2 + x \to$ neither

f is neither

c) $f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x - y \\ \pi x + 6y \\ 3y + 1 \end{bmatrix}$ $\begin{matrix} \to \text{linear} \\ \to \text{linear} \\ \to \text{affine} \end{matrix}$

f is affine

only linear if $\vec{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$\uparrow\ \vec{b} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Note that a general linear function $f : \mathbb{R}^3 \to \mathbb{R}^3$ looks like

$$f\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} ax + by + cz \\ dx + ey + gz \\ hx + iy + jz \end{bmatrix},$$

which have a lot of constants to keep track of ($9 = 3 \cdot 3$ in fact), and there could be a lot more for a general $f : \mathbb{R}^n \to \mathbb{R}^m$ (mn to be exact). We introduce a shorthand notation.

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

m x n
↳ m rows, n columns

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

1st number = row
2nd number = column

- The collection of entries $\begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,n} \end{bmatrix}$ along the ith horizontal row (with $i = 1$ along the top side) is called the **ith row**, and the collection of entries $\begin{bmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{bmatrix}$ along the jth vertical layer (with $j = 1$ along the left side) is called the **jth column**.
- The entry in row i and column j, $a_{i,j}$, is called the ij**-entry** or (i,j)**-entry**.

We have a notion of multiplication between a matrix and a vector:

Definition: 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

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

dot product
– each row by the vector column

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

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

$$\to x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

Note: You need to pay attention to all of the dimensions going on; this product is only defined for an $m \times n$ matrix multiplied by an n-vector. It produces an m-vector.

$$\underset{\substack{\uparrow \\ m \times n \text{ matrix}}}{(m \times n)} \; \underset{\substack{\uparrow \\ n\text{-vector}}}{(n \times 1)} = \underset{\substack{\uparrow \\ m\text{-vector}}}{(m \times 1)}$$

Example 2: Compute the following matrix-vector products.

(a) $\begin{bmatrix} 2 & -1 & 3 \\ 4 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2(1) + (-1)(2) + (3)(3) \\ 4(1) + (1)(2) + (0)(2) \\ 0(1) + (-1)(2) + (1)(3) \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 1 \end{bmatrix}$

(b) $\begin{bmatrix} -4 & 8 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} (-4)(1) + 8(0) \\ (1)(1) + 2(0) \end{bmatrix} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$

★ note: multiplying by $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ gives us the 1^{st} column

$f: n \to m$, matrix is $m \times n$

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

A consequence of this proposition is that an affine function $f : \mathbb{R}^n \to \mathbb{R}^m$ can be written as $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, where A is an $m \times n$ matrix, $\mathbf{x}$ and n-vector, and $\mathbf{b}$ an m-vector.

Example 3: Write the following linear/affine functions in the form $A\mathbf{x} + \mathbf{b}$.

(a) $f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x - y \\ \pi x + 6y \\ 3y \end{bmatrix} = \overset{A}{\begin{bmatrix} 2 & -1 \\ \pi & 6 \\ 0 & 3 \end{bmatrix}} \overset{\vec{x}}{\begin{bmatrix} x \\ y \end{bmatrix}} = \begin{bmatrix} 2x - y \\ \pi x + 6y \\ 0x + 3y \end{bmatrix}$

linear

$f: \mathbb{R}^2 \to \mathbb{R}^3$, matrix should be 3×2

(b) $f\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 2x + z + 4 \\ x + y + z \\ -x - 2y + 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2x + z \\ x + y + z \\ -x - 2y \\ 0 \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \\ 1 \\ 3 \end{bmatrix} = \overset{A}{\begin{bmatrix} 2 & 0 & 1 \\ 1 & 1 & 1 \\ -1 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix}} \overset{\vec{x}}{\begin{bmatrix} x \\ y \\ z \end{bmatrix}} + \overset{\vec{b}}{\begin{bmatrix} 4 \\ 0 \\ 1 \\ 3 \end{bmatrix}}$

affine

$f: \mathbb{R}^3 \to \mathbb{R}^4$, matrix should be 4×3

each of these are vectors

Theorem 13.4.1: If $\mathbf{c}_1, \ldots \mathbf{c}_n$ are the columns of A, i.e. $A = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \cdots \quad \mathbf{c}_n]$ then

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

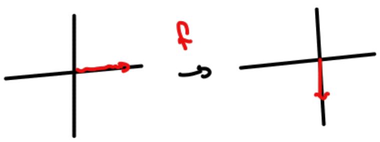
In other words, the matrix-vector product is just a linear combination of the columns of A, where the coefficients are the entries of the vector $\mathbf{x}$.

Theorem 13.4.5: For a linear function $f(\mathbf{x}) = A\mathbf{x}$, the matrix A has its respective columns $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)$, where $\mathbf{e}_i$ is the ith standard basis vector in $\mathbb{R}^n$. This gives us a way to reconstruct the matrix A given we know f.

$\longrightarrow f(\vec{x}) = A\vec{x} = \left[f(\vec{e}_1) \cdots f(\vec{e}_n) \right] \vec{x}$

$f(\vec{e}_n)$ lists e_n as a column

Example 4: Recall the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ which rotated vectors $90°$ clockwise.



$$f(\vec{e}_1) = f\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$f(\vec{e}_2) = f\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Using theorem 13.4.5, $f\begin{pmatrix} x \\ y \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ -x \end{bmatrix}$

$$f\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{array}{cc} c_1 & c_2 \\ \multicolumn{2}{c}{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}} \end{array} \begin{bmatrix} 2 \\ 3 \end{bmatrix}\begin{array}{l} \leftarrow x_1 \\ \leftarrow x_2 \end{array} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \overset{x_1}{2}\overset{c_1}{\begin{bmatrix} 0 \\ -1 \end{bmatrix}} + \overset{x_2}{3}\overset{c_2}{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}$$

$$= \begin{bmatrix} 0 \\ -2 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Theorem 13.4.1

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

component functions

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

We also refer to $Df(\mathbf{a})$ as the **Jacobian matrix** of f at the point $\mathbf{a}$. In general, the ith row of $Df(\mathbf{a})$ is $\nabla f_i(\mathbf{a})$ written horizontally.

the i^{th} row is the gradient of the i^{th} component function of the vector-valued function f

If f is not a linear, can we approximate it via $f(\mathbf{x}) \approx f(\mathbf{a}) + L(\mathbf{x}-\mathbf{a})$, where L is linear? If so, what is the "best" one? Here is the answer:

<u>Theorem 13.5.8</u>: The best linear approximation to $f : \mathbb{R}^n \to \mathbb{R}^m$ at the point $\mathbf{a}$ is given by the derivative matrix $Df(\mathbf{a})$. We have

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \underbrace{Df(\mathbf{a})(\mathbf{x}-\mathbf{a})}_{}$$

matrix vector product

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

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

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

Example 5: Work out the linear approximations $f(\mathbf{a})$ and $f(\mathbf{a}+\mathbf{h})$, where $\mathbf{a} = (1,1,1)$, for the function f below. Then, estimate $f(.9, 1.1, 1.2)$.

$$f(x,y,z) = \begin{bmatrix} x^2+yz \\ xyz \\ \sqrt{xz} \end{bmatrix}. \begin{matrix} \leftarrow f_1 \\ \leftarrow f_2 \\ \leftarrow f_3 \end{matrix}$$

Compute $Df(\vec{x})$

$$\nabla f_1 = \begin{bmatrix} 2x \\ z \\ y \end{bmatrix} \rightarrow 1^{st} \text{ row}$$

$$\nabla f_2 = \begin{bmatrix} yz \\ xz \\ xy \end{bmatrix} \rightarrow 2^{nd} \text{ row}$$

$$\frac{z}{\sqrt{xz}} \cdot \frac{\sqrt{z}}{\sqrt{z}} = \frac{z\sqrt{z}}{z\sqrt{x}} = \frac{\sqrt{z}}{\sqrt{x}} = \sqrt{\frac{z}{x}}$$

$$\nabla f_3 = \begin{bmatrix} \frac{1}{2}\frac{z}{\sqrt{xz}} \\ 0 \\ \frac{1}{2}\frac{x}{\sqrt{xz}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{\frac{z}{x}} \\ 0 \\ \frac{1}{2}\sqrt{\frac{x}{z}} \end{bmatrix}$$

$$Df(\vec{x}) = \begin{bmatrix} 2x & z & y \\ yz & xz & xy \\ \frac{1}{2}\sqrt{\frac{z}{x}} & 0 & \frac{1}{2}\sqrt{\frac{x}{z}} \end{bmatrix}, \quad Df(1,1,1) = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$f(\vec{x}) \approx f\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1/2 & 0 & 1/2 \end{pmatrix}\left(\vec{x} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right)$$

$$\approx \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1/2 & 0 & 1/2 \end{bmatrix}\begin{bmatrix} x-1 \\ y-1 \\ z-1 \end{bmatrix}$$

$$f(\vec{a}+\vec{h}) \approx f(\vec{a}) + Df(\vec{a})\vec{h}$$

$$\approx \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1/2 & 0 & 1/2 \end{bmatrix} \vec{h}$$

$$f(\vec{x}) \approx \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} x-1 \\ y-1 \\ z-1 \end{bmatrix}$$

$$f\begin{pmatrix} 0.9 \\ 1.1 \\ 1.2 \end{pmatrix} \approx \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 0.9-1 \\ 1.1-1 \\ 1.2-1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} -0.1 \\ 0.1 \\ 0.2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -0.2 + 0.1 + 0.2 \\ -0.1 + 0.1 + 0.2 \\ -0.05 + 0 + 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.05 \end{bmatrix}$$

$$f\begin{pmatrix} 0.9 \\ 1.1 \\ 1.2 \end{pmatrix} \approx \begin{bmatrix} 2.1 \\ 1.2 \\ 1.05 \end{bmatrix}$$

Actual Answer:

$$f\begin{pmatrix} 0.9 \\ 1.1 \\ 1.2 \end{pmatrix} = \begin{bmatrix} 2.13 \\ 1.188 \\ 1.04 \end{bmatrix}$$

it's a pretty good approximation