Pre-Reading due Monday:

Lecture 11 - Gradients, Local Approximations, and Gradient Descent

October 21, 2022

Goals: Compute the gradient vector of a function $f : \mathbb{R}^n \to \mathbb{R}$, then use it to compute local approximations of f, the tangent plane to a level set, and identify the directions of most rapid increase/decrease.

Definition: Let $f : \mathbb{R}^n \to \mathbb{R}$. The **gradient of** f is defined to be

$$\nabla f = \begin{bmatrix} f_{x_1} \\ \vdots \\ f_{x_n} \end{bmatrix}.$$

$\nabla f : \mathbb{R}^n \to \mathbb{R}^n$

Note that the gradient of f is a vector-valued function $\mathbb{R}^n \to \mathbb{R}$; its value $(\nabla f)(\mathbf{a})$ is an n-vector. For $\mathbf{x}$ near $\mathbf{a} \in \mathbb{R}^n$, the **linear approximation** to f is

$$f(\vec{\mathbf{x}}) \approx f(\vec{\mathbf{a}}) + \nabla f(\vec{\mathbf{a}}) \cdot (\vec{\mathbf{x}} - \vec{\mathbf{a}}).$$

For $f : \mathbb{R}^2 \to \mathbb{R}$, we have in particular that

change in x

↳ dot product bc things are vectors

change in y

$$\boxed{f(x,y) \approx f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b).}$$

first order approximation

Example 1: Compute the gradient of the function $f(x,y) = x^2y^2 - 2x^4y$. Use linear approximation to estimate $f(.8, 1.1)$.

$f_x = 2xy^2 - 8x^3y$

$f_y = 2x^2y - 2x^4$

$$\nabla f(x,y) = \begin{bmatrix} 2xy^2 - 8x^3y \\ 2x^2y - 2x^4 \end{bmatrix}$$

approximate $f(.8, 1.1) \to$ use $f(1,1)$

$f(1,1) = 1 - 2 = -1$, $\quad \nabla f(1,1) = \begin{bmatrix} 2-8 = -6 \\ 2-2=0 \end{bmatrix}$

$f(.8, 1.1) = f(1,1) + f_x(1,1)(.8-1) + f_y(1,1)(1.1-1)$

$= -1 + (-6)(-.2) + (0)(.1)$

$= -1 + 1.2 = \boxed{.2}$

$= -1 +$

Theorem 11.2.1: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a scalar-valued function, and suppose $\nabla f(a,b) \neq \mathbf{0}$.

- The gradient $\nabla f(a,b)$ is perpendicular to the level set of f that goes through (a,b) (i.e. it is perpendicular to the tangent line of the level set through (a,b)). It **points in the direction of maximal increase** for $f(x,y)$ for (x,y) moving away from (a,b).
- The equation

$$\nabla f(a,b) \cdot \begin{bmatrix} x-a \\ y-b \end{bmatrix} = 0$$

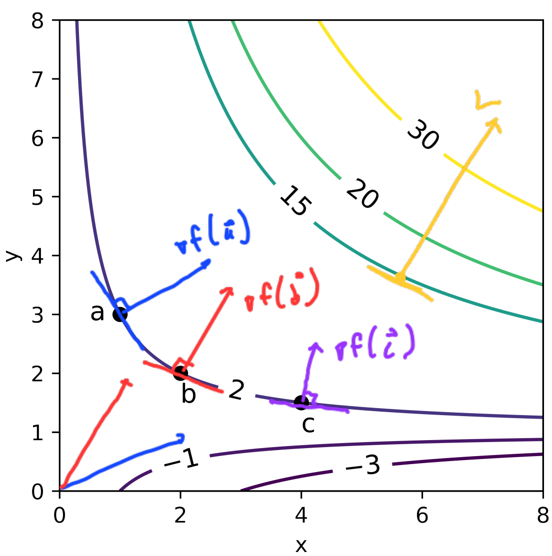
bc it's perpendicular to tangent line

in the (x,y)-plane is the line tangent to the curve of the contour plot of $f(x,y)$ through $(x,y) = (a,b)$. Explicitly, the equation of this line is

$$\boxed{f_x(a,b)(x-a) + f_y(a,b)(y-b) = 0.}$$

Example 2: Here is the contour plot of the function $f(x,y) = xy - x$.

$\vec a = (1,3)$, $\vec b = (2,2)$, $\vec c = (4, \frac{3}{2})$ — level curve $f(x,y) = xy - x = 2$



$\nabla f(x,y) = \begin{bmatrix} y-1 \\ x \end{bmatrix}$

$\nabla f(\vec a) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\nabla f(\vec b) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$\nabla f(\vec c) = \begin{bmatrix} 1/2 \\ 4 \end{bmatrix}$

Tangent line @ $(1,3) = \vec a$

$2(x-1) + 1(y-3) = 0$ using eqn above

$y = -2x+5 \to \begin{bmatrix} x \\ -2x+5 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \end{bmatrix} + x\begin{bmatrix} 1 \\ -2 \end{bmatrix}$

check: $\begin{pmatrix} 2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2 \end{pmatrix} = 0$ ✓

$\nabla f(\vec a)$

vector of the line

TL @ $\vec b = (2,2)$

$1\cdot(x-2) + 2(y-2) = 0$

TL @ $\vec c = (4, 3/2)$

$\frac{1}{2}(x-4) + 4(y - \frac{3}{2}) = 0$

Example 3: Find the parametric form of the line that intersects the curve $xy - x = 15$ at the point $(5,4)$ perpendicularly.

$\begin{bmatrix} x-a \\ y-b \end{bmatrix} = \begin{bmatrix} x-1 \\ y-3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \to \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

Line: $\begin{pmatrix} 5 \\ 4 \end{pmatrix} + t\vec v$

$\vec v = \nabla f\begin{pmatrix} 5 \\ 4 \end{pmatrix} = \begin{pmatrix} y-1 \\ x \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$

Line: $\begin{pmatrix} 5 \\ 4 \end{pmatrix} + t\begin{pmatrix} 3 \\ 5 \end{pmatrix}$, $t \in \mathbb{R}$

Theorem 11.2.2: For a scalar-valued function $f : \mathbb{R}^3 \to \mathbb{R}$ and a point $\mathbf{a}$ for which $\nabla f(\mathbf{a}) \neq 0$, the gradient vector is perpendicular to the *plane tangent* to the level set of f through $\mathbf{a}$. This **tangent plane** has the equation

tangent surface

generally not a subspace, won't include origin

$$\underbrace{\nabla f(a_1, a_2, a_3)}_{\text{normal}} \cdot \overbrace{\begin{bmatrix} x - a_1 \\ y - a_2 \\ z - a_3 \end{bmatrix}}^{\text{displacement vector to } \vec{a}} = 0.$$

$\to f_x(\vec{a})(x - a_1) + f_y(\vec{a})(y - a_2) + f_z(\vec{a})(z - a_3) = 0$

Example 3: The graph of a function $h : \mathbb{R}^2 \to \mathbb{R}$ is the surface S with equation $z = h(x, y)$ that is the level set $f = 0$ of the function $f(x, y, z) = z - h(x, y)$ whose gradient $(-h_x, -h_y, 1)$ never vanishes.

$\hookrightarrow \nabla f = (-h_x, -h_y, 1)$

Example 4: Find the tangent plane to the surface $z = x^2 + y^2$ at the point $(1, 1, 2)$.

$\hookrightarrow f = 0$ level set

gradient vector is normal of tangent plane

$$f(x, y, z) = z - (x^2 + y^2)$$

$$\nabla f = \begin{pmatrix} -2x \\ -2y \\ 1 \end{pmatrix} \to \nabla f(1, 1, 2) = \begin{pmatrix} -2 \\ -2 \\ 1 \end{pmatrix}$$

Tangent Plane

let $\vec{w} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ on plane

$\vec{n} \cdot (\vec{w} - \vec{p}) = 0$

$$\begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} x - 1 \\ y - 1 \\ z - 2 \end{bmatrix} = 0 \to -2(x - 1) - 2(y - 1) + (z - 2) = 0$$

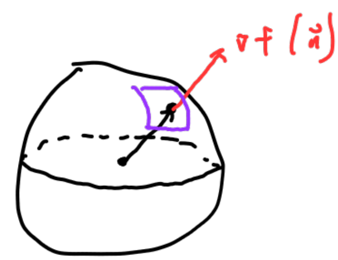
$$\boxed{-2x - 2y + z = -2}$$

Example 5: Find the equation of the tangent plane to the unit sphere at the point $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$.

unit sphere: $x^2 + y^2 + z^2 = 1$

$f(x, y, z) = x^2 + y^2 + z^2$

$\hookrightarrow$ unit sphere is $f = 1$ level set



$\vec{n}$ = the point itself

$$\nabla f = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix} \to \nabla f(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}) = \begin{pmatrix} 2/\sqrt{3} \\ 2/\sqrt{3} \\ 2/\sqrt{3} \end{pmatrix}$$

$x^2 + y^2 + z^2 = r^2$

$\vec{n} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = r^2$

let $(a, b, c) = (x, y, z)$

$\vec{n} \cdot \begin{pmatrix} x - a \\ y - b \\ z - c \end{pmatrix} = 0 \to$ constant on RHS will be r^2

Tangent plane

$$\begin{pmatrix} 2/\sqrt{3} \\ 2/\sqrt{3} \\ 2/\sqrt{3} \end{pmatrix} \cdot \begin{pmatrix} x - 1/\sqrt{3} \\ y - 1/\sqrt{3} \\ z - 1/\sqrt{3} \end{pmatrix} = 0 \to \frac{1}{\sqrt{3}} x + \frac{1}{\sqrt{3}} y + \frac{1}{\sqrt{3}} z = 1$$

Theorem 11.3.2: Let $f : \mathbb{R}^n \to \mathbb{R}$ be a scalar-valued function and $\mathbf{a} \in \mathbb{R}^n$ a point at which the gradient $\nabla f(\mathbf{a}) \neq \mathbf{0}$. The unit vector $\nabla f(\mathbf{a})/\|\nabla f(\mathbf{a})\|$ is the direction in which f increases most rapidly at $\mathbf{a}$. Likewise, the opposite unit vector $-\nabla f(\mathbf{a})/\|\nabla f(\mathbf{a})\|$ is the direction in which f decreases most rapidly at $\mathbf{a}$.

moving this way leads to a minimum

Example 6: Gradient descent is an algorithm for finding a minimum of a function using the result of the above theorem. We will walk through the first few steps of gradient descent for the function

$$z = f(x, y) = x^2 + y^2.$$

The algorithm: start at a point $\mathbf{a}$, then move to the point $\mathbf{a}_1 = \mathbf{a} + t\nabla f(\mathbf{a})$. Next, move to $\mathbf{a}_2 = \mathbf{a}_1 + t\nabla f(\mathbf{a}_1)$. Repeat until the minimum is reached.

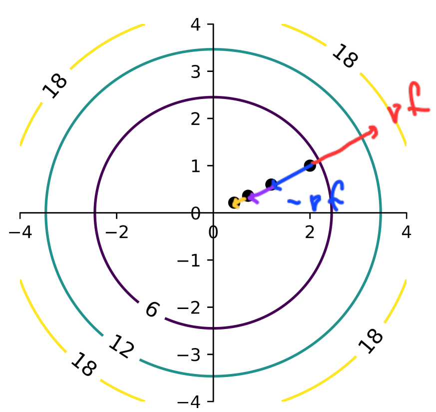
So we need two things:

learning rate

1. A starting point $\mathbf{a}$.
2. t - how far to go at each step. In machine learning, this is called the **learning rate**. We want $t < 0$ since then $t\nabla f(\mathbf{a})$ will point in the direction of maximal decrease.

Here are the first 3 steps:

$\nabla f = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$



$\vec{a} = (2,1)$, $t = -0.2$

$\vec{a}_1 = \vec{a} + t\nabla f(\vec{a})$

$= \begin{pmatrix} 2 \\ 1 \end{pmatrix} - 0.2\, \nabla f \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - 0.2 \begin{pmatrix} 4 \\ 2 \end{pmatrix}$

$= \begin{pmatrix} 1.2 \\ 0.6 \end{pmatrix}$

$\vec{a}_2 = \vec{a}_1 + t\nabla f(\vec{a}_1)$

$= \begin{pmatrix} 1.2 \\ 0.6 \end{pmatrix} - 0.2 \begin{pmatrix} 2.4 \\ 1.2 \end{pmatrix} = \begin{pmatrix} 0.72 \\ 0.36 \end{pmatrix}$

$\vec{a}_3 = \begin{pmatrix} .72 \\ .36 \end{pmatrix} - 0.2 \begin{pmatrix} 1.44 \\ .72 \end{pmatrix} = \begin{pmatrix} .432 \\ .216 \end{pmatrix}$

Note: The 18th step gives our global minimum (0,0) when rounded to 3 decimals. The 12th step gives (.004,.002) when rounded, which is pretty close to (0,0), so it might not be worth the trouble to do extra steps for such little gain, especially if we are dealing with much more complicated functions.

↳ Another way: $X, Y \to$ is our data

squared error: $\sum_{i=1}^{n} \left(y_i - (m x_i + b)\right)^2$

$= f(m, b)$

$$f_m = \sum_{i=1}^{n} 2 \left(y_i - (m x_i + b) \right) (-x_i)$$

$$= \sum_{i=1}^{n} -2 \left(y_i - (m x_i + b) \right) x_i$$

$$f_b = \sum_{i=1}^{n} 2 \left(y_i - (m x_i + b) \right) (-1)$$

$$= \sum_{i=1}^{n} -2 \left(y_i - (m x_i + b) \right)$$

You can use these functions w/ gradient descent to find line of best fit.