Gradient Vector

Thursday, 20 March 2025 12:19 pm

α

$=$

Δ δ

For a scalar-valued function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient of f is defined as:

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

The gradient is a vector-valued function $\mathbb{R}^n \to \mathbb{R}^n$: its value $(\nabla f)(\mathbf{a})$ at $\mathbf{a} \in \mathbb{R}^n$ is an n-vector.

2-Vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

3-Vec $\begin{bmatrix} 0.5 \\ -1 \\ 0 \end{bmatrix}$ n-ve

R^2

$$f(x, y) = 3x^2y^2 - 5xy^3$$

Find ∇f at

Point $(-1, 0)$

$$\nabla f = \begin{bmatrix} 6xy^2 - 5y^3 \\ 6x^2y - 15xy^2 \end{bmatrix}$$

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

Example

Thursday, 20 March 2025 12:20 pm

Find the gradient vector of $f(x,y) = \underline{x}^2 + \underline{2y}^2 + \underline{xy}$ at point (2,3)

$$\nabla f = \begin{bmatrix} 2x + y \\ 4y + x \end{bmatrix}$$

$$\nabla f(\overset{x}{2}, \overset{y}{3}) = \begin{bmatrix} 7 \\ 14 \end{bmatrix} = 7 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Problem 1

Thursday, 20 March 2025 12:24 pm

Calculate ∇f where $f(x, y, z) = x^2yz + y^2z + xyz^2$ at the point (2,1,3).

$$\nabla f(2,1,3) = \begin{bmatrix} 21 \\ 36 \\ 17 \end{bmatrix} \begin{array}{l} \rightarrow \partial f/\partial x \\ \rightarrow \partial f/\partial y \\ \rightarrow \partial f/\partial z \end{array}$$

Problem 2

Thursday, 20 March 2025 12:25 pm

For the scalar field $\varphi(x, y) = ln(x^2 + y^2)$, find the gradient vector at the point (3,4).

$$\nabla f(3,4) = \begin{bmatrix} 6/25 \\ 8/25 \end{bmatrix} \qquad \nabla f = \begin{bmatrix} \frac{1}{x^2+y^2}(2x) \\ \frac{1}{x^2+y^2}(2y) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{2x}{x^2+y^2} \\ \frac{2y}{x^2+y^2} \end{bmatrix}$$

$$df = \frac{df}{dx} = \frac{\overset{y_2 - y_1}{f(x+h) - f(x)}}{\underset{x_2 - x_1}{(x+h) - (x)}}$$

$$df(x_2 - x_1) = y_2 - y_1$$

$$y_2 = y_1 + df \cdot (x_2 - x_1)$$

$$f(x_2) = f(x_1) + df \cdot (x_2 - x_1)$$

$$\boxed{f(x) = f(x_1) + df \cdot (x - x_1)}$$

$$f(x) = \overbrace{f(x_1)}^{b} + \overbrace{\nabla f(x_1)}^{m}\overbrace{(x - x_1)}^{x}$$

$$f(X) = f(X_1) + \nabla f(X_1)(X - X_1)$$

Exp

$$f(x,y) = \underbrace{x^2+y^2}$$

$(1,1)$

$$X = \begin{bmatrix} x \\ y \end{bmatrix} \longleftrightarrow X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$f(X_1) = 2$

$$\nabla f = \begin{bmatrix} 2x \\ 2y \end{bmatrix} \qquad \nabla f(X_1) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$f(X) = 2 + \begin{bmatrix} 2 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} x-1 \\ y-1 \end{bmatrix}$$

$$= 2 + 2(x-1) + 2(y-1)$$

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

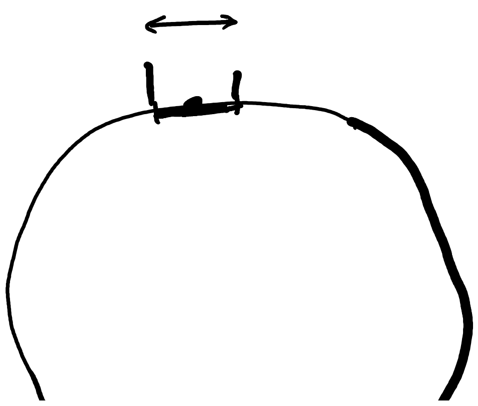
$$f(x,y) = x^2+y^2 \approx 2x+2y-2$$

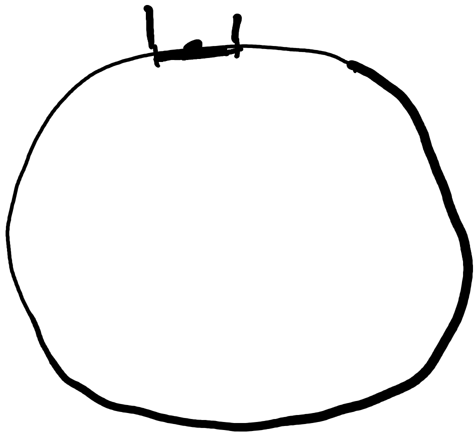
at $(1,1)$

$2.65 \longleftrightarrow 2.6$

$(1.1,\ 1.2)$

$(1,1)$





Linear Approximation using Gradients

Thursday, 20 March 2025 12:29 pm

For a function $f : \mathbb{R}^n \to \mathbb{R}$ and a point $\mathbf{a} \in \mathbb{R}^n$, the linear approximation to f near $\mathbf{a}$ is:

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

Example 2

Thursday, 20 March 2025 12:31 pm

Find the linear approximation of $f(x, y) = x^2 + 2y^2 + xy$ at point (1,1)

$$f(1,1) = 1 + 2(1) + (1)(1)$$

$$= 1 + 2 + 1 = 4$$

$$\frac{\partial f}{\partial x} = 2x + y$$

$$\frac{\partial f}{\partial y} = 4y + x$$

$$\frac{\partial f}{\partial y} = 4(1) + 1 = 5$$

$$\frac{\partial f}{\partial x} = 2(1) + 1 = 3$$

$$\nabla f(1,1) = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

$$f(x) = f(x_1) + \nabla f(1,1)(x - x_1)$$

$$= 4 + \begin{bmatrix} 3 \\ 5 \end{bmatrix} \begin{bmatrix} x-1 \\ y-1 \end{bmatrix}$$

$$= 4 + 3(x-1) + 5(y-1)$$

$$4 + 3x - 3 + 5y - 5 = 3x + 5y - 4$$

Problem 3

Thursday, 20 March 2025 12:31 pm

For $f(x,y) = \sqrt{x^2+y^2}$, compute the linear approximation of f at the point $(3,4)$. Use this approximation to estimate $f(3.1, 3.9)$.

$$f(x,y) = \sqrt{x^2+y^2} \longrightarrow (3,4)$$

$$\boxed{(3.1,\ 3.9)}$$

$$\tilde{f}(x,y) = 1.2x + 1.6y - \cancel{8} \ ? \qquad 0.6x + 0.8y$$

$$3x + 4y \ \times \ ?$$

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(x^2+y^2)^{1/2} = \frac{1}{\cancel{2}}(x^2+y^2)^{-1/2}(\cancel{2}x)$$

$$= \frac{x}{\sqrt{x^2+y^2}} \qquad \frac{\partial f}{\partial y} = \frac{y}{\sqrt{x^2+y^2}}$$

$$\nabla f = \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2}} \\ \frac{y}{\sqrt{x^2+y^2}} \end{bmatrix} \qquad \nabla f(3,4) = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix}$$

$$\tilde{f}(x,y) = \overset{5}{f(3,4)} + \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix} \begin{bmatrix} x-3 \\ y-4 \end{bmatrix}$$

$$= 5 + \frac{3}{5}(x-3) + \frac{4}{5}(y-4)$$

$$= 5 - \frac{9}{5} - \frac{16}{5} + \frac{3}{5}x + \frac{4}{5}y$$

$$= \frac{25-9-16}{5} + \cancel{0.6x + 0.8y}$$

$$= \frac{25 - y - 16}{5} + \boxed{0.6x + 0.8y}$$

Problem 4

Thursday, 20 March 2025 12:42 pm

For $f(x, y) = \ln(x^2 + y^2)$, compute the linear approximation of f at the point $(1, 2)$. Use this approximation to estimate $f(0.9, 2.1)$.

Gradient Descent

Thursday, 20 March 2025 12:43 pm

Algorithm: 1. Start at an initial point $\mathbf{x}_0$. 2. Update: $\mathbf{x}_{k+1} = \mathbf{x}_k - t \cdot \nabla f(\mathbf{x}_k)$, where $t > 0$ is a small step size. 3. Repeat step 2 until convergence or a maximum number of iterations is reached.

The parameter t is often called the learning rate in machine learning contexts.

To find X_{opt} where function $f(X)$ is minimum, we use

$$\underbrace{X_{k+1}}_{X} = \underbrace{X_k}_{X} - \alpha \nabla f(X_k)$$

$$X = X - \alpha \nabla$$

$$f(x, y) = x^2 + 2xy + y^2 \qquad X_{opt} = \begin{bmatrix} x \\ y \end{bmatrix} \begin{matrix} ? \\ ? \end{matrix}$$

$$\nabla f = \begin{bmatrix} 2x + 2y \\ 2x + 2y \end{bmatrix} \qquad \alpha = 0.1 \qquad f = \min \downarrow$$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

X_k	$\nabla f(X_k)$	$X_{k+1} = X_k - \alpha \nabla f(X_k)$
$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ infeasible initial pt.
$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 4 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0.1 \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix}$
$\begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix}$	$\begin{bmatrix} 2.4 \\ 2.4 \end{bmatrix}$	$\begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix} - 0.1 \begin{bmatrix} 2.4 \\ 2.4 \end{bmatrix} = \begin{bmatrix} 0.36 \\ 0.36 \end{bmatrix}$

$$\nabla f = \begin{bmatrix} 2x + 2y \\ 2x + 2y \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\nabla f = AX$$

$$\nabla f = AX$$

Example 3

Thursday, 20 March 2025 12:48 pm

Let's use gradient descent to find a minimum of the function:

$$f(x, y) = x^2 - 3xy + 3y^2 + 5y + 2x$$

$$\nabla f = \begin{bmatrix} 2x - 3y + 2 \\ -3x + 6y + 5 \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ -3 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$A \quad X \quad + \quad B$$

$$\alpha = 0.1$$

$$X_{k+1} = X_k - \alpha \nabla f(X_k)$$

X_k	$\nabla f(X_k)$	$X_{k+1} = X_k - \alpha \nabla f(X_k)$
$(0,0)$	$(2,5)$	$X_{k+1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} -0.2 \\ -0.5 \end{bmatrix}$
$\begin{bmatrix} -0.2 \\ -0.5 \end{bmatrix}$	$\begin{bmatrix} 3.1 \\ 2.6 \end{bmatrix}$	$\begin{bmatrix} -0.2 \\ -0.5 \end{bmatrix} - 0.1 \begin{bmatrix} 3.1 \\ 2.6 \end{bmatrix} = \begin{bmatrix} -0.51 \\ -0.76 \end{bmatrix}$

$f(x, y)$

$y \qquad x$

$$\underline{J(w_0, w_1)} = \frac{1}{2m} \sum_{i=1}^{m} (w_0 + w_1 x - y)^2$$

$$\nabla J = \begin{bmatrix} \frac{\partial J}{\partial w_0} \\ \frac{\partial J}{\partial w_1} \end{bmatrix}$$

$$W = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \qquad W = W - \alpha \nabla J$$

Problem 5

Thursday, 20 March 2025 12:52 pm

Consider the function $f(x, y) = x^2 + 4y^2$. Starting from the point $(3, 2)$, perform two iterations of gradient descent using a step size of $t = 0.1$.

Problem 6

Thursday, 20 March 2025 12:53 pm

For the function $f(x, y) = (x-3)^2 + (y+2)^2$, explain where gradient descent will converge starting from any initial point, and why.