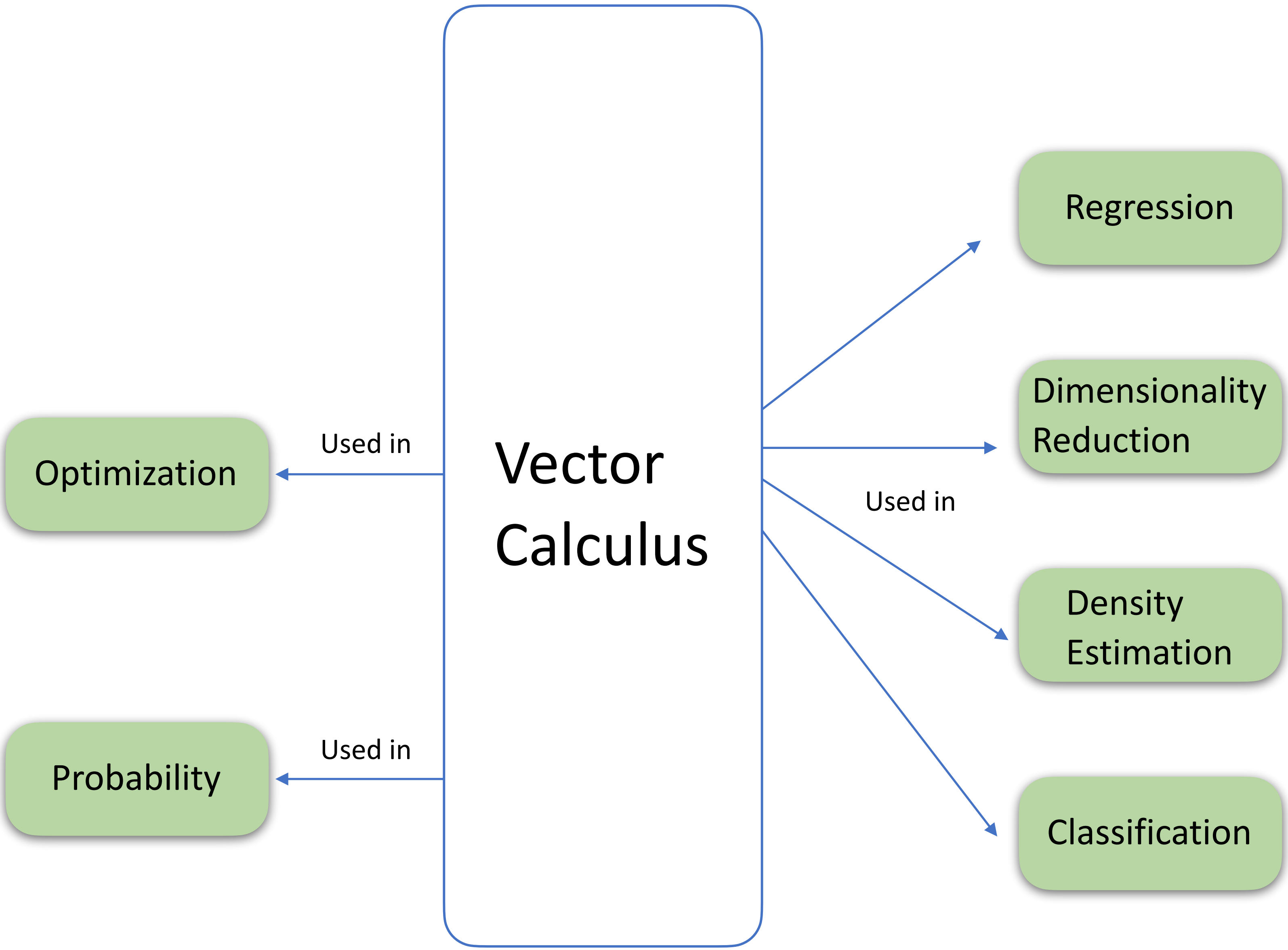
Vector Calculus I

Jo Ciucă
Australian National University
comp36706670@anu.edu.au


Vector Calculus
Used in
Optimization
Used in
Probability
Used in
Regression
Dimensionality Reduction
Density Estimation
Classification

Outline

- Differentiation of Univariate Functions
- Partial Differentiation
- Gradients

5 Vector Calculus

- We discuss functions

$$f : \mathbb{R}^D \to \mathbb{R}$$

$$\boldsymbol{x} \mapsto f(\boldsymbol{x})$$

where $\mathbb{R}^D$ is the domain of f, and the function values $f(\boldsymbol{x})$ are the image/codomain of f.

- Example (dot product)
- Previously, we write dot product as:

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}, \quad \boldsymbol{x} \in \mathbb{R}^2$$

- In this lecture, we write it as:

$$f : \mathbb{R}^2 \to \mathbb{R}$$

$$\boldsymbol{x} \mapsto x_1^2 + x_2^2$$

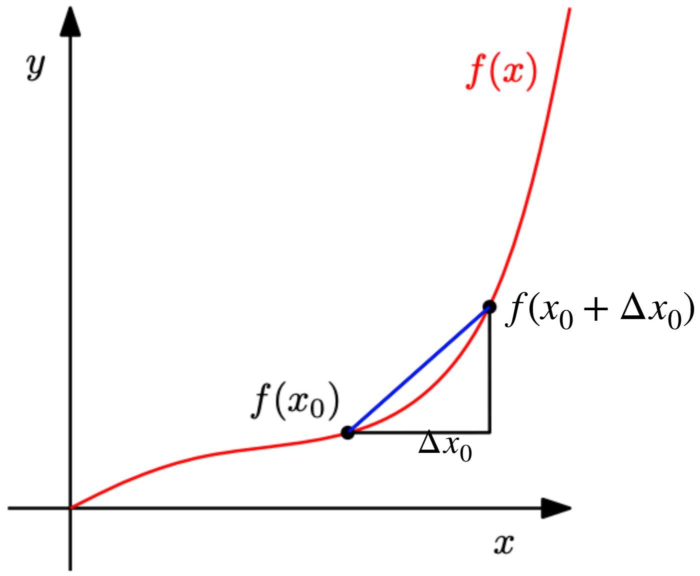
5.1 Differentiation of Univariate Functions

- Given $y = f(x)$, the difference quotient is defined as

$$\frac{\Delta y}{\Delta x} := \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

- It computes the slope of the secant line through two points on the graph of f. In this figure, these are the points with x–coordinates x_0 and $x_0 + \Delta x_0$.

- In the limit for $\Delta x \to 0$, we obtain the tangent of f at x (if f is differentiable). The tangent is then the derivative of f at x.

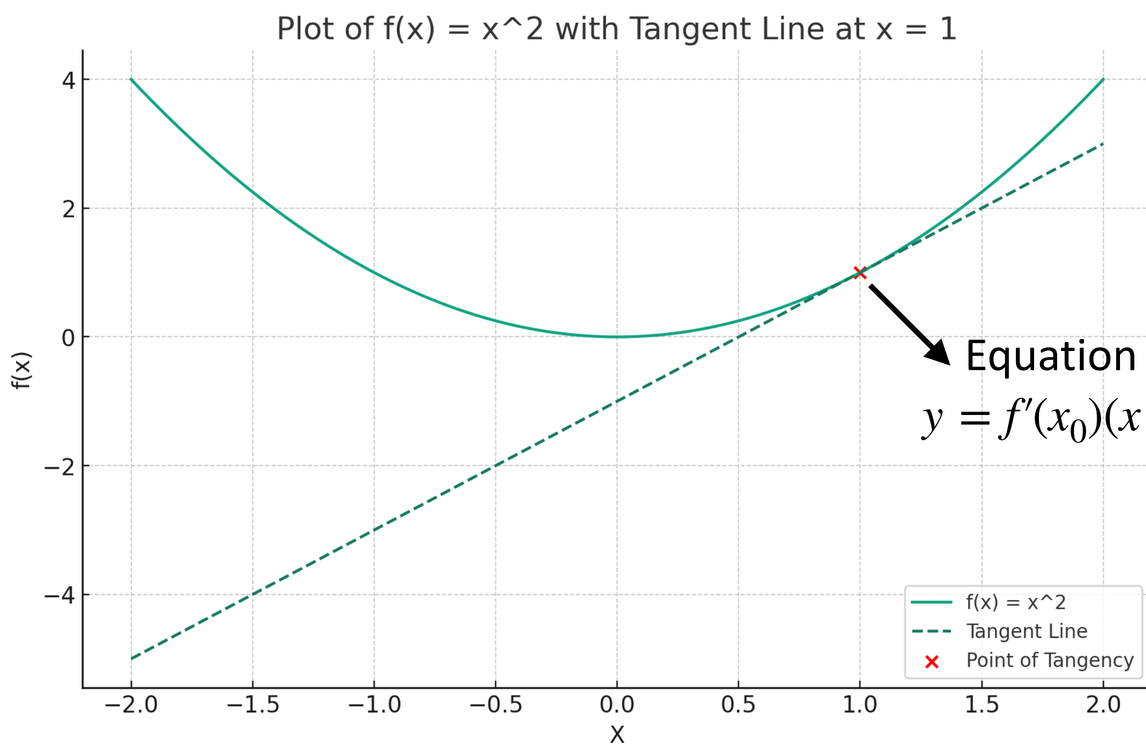


5.1 Differentiation of Univariate Functions

- For $h > 0$, the derivative of f at x is defined as the limit

$$\frac{\mathrm{d}f}{\mathrm{d}x} := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \qquad = f'(x)$$

- The derivative of f points in the direction of steepest ascent of f.



Equation of the line

$$y = f'(x_0)(x - x_0) + f(x_0)$$

5.1 Differentiation of Univariate Functions

- Example - Derivative of a Polynomial
- Compute the derivative of $f(x) = x^n, \quad n \in \mathbb{N}$. (From our high school knowledge, the derivative is nx^{n-1}.)

$$\frac{\mathrm{d}f}{\mathrm{d}x} := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$$= \lim_{h \to 0} \frac{(x+h)^n - x^n}{h}$$

$$= \lim_{h \to 0} \frac{\sum_{i=0}^{n} \binom{n}{i} x^{n-i} h^i - x^n}{h}$$

we see that $x^n = \binom{n}{0} x^{n-0} h^0$. By starting the sum at 1, the x^n cancels.

5.1 Differentiation of Univariate Functions

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{h\to 0} \frac{\sum_{i=0}^{n}\binom{n}{i}x^{n-i}h^{i} - x^{n}}{h}$$

$$= \lim_{h\to 0} \frac{\sum_{i=1}^{n}\binom{n}{i}x^{n-i}h^{i}}{h}$$

$$= \lim_{h\to 0} \sum_{i=1}^{n}\binom{n}{i}x^{n-i}h^{i-1}$$

$$= \lim_{h\to 0} \left\{ \binom{n}{1}x^{n-1} + \underbrace{\sum_{i=2}^{n}\binom{n}{i}x^{n-i}h^{i-1}}_{\to 0 \; as \; h \to 0} \right\}$$

$$= nx^{n-1}$$

5.1.2 Differentiation Rules

- Product rule

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$$

- Quotient rule:

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$$

- Sum rule:

$$(f(x) + g(x))' = f'(x) + g'(x)$$

- Chain rule:

$$(g(f(x)))' = (g \circ f)'(x) = g'(f(x))f'(x)$$

Here, $g \circ f$ denotes function composition $g\big(f(x)\big)$

Tools in our maths toolbox

Example: Chain rule

- Compute the derivative of the function $h(x) = (2x+1)^4$
- We write

$$h(x) = (2x+1)^4 = g\left(f(x)\right)$$

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

$$g\left(f\right) = f^4$$

- We obtain the derivatives of f and g as,

$$f'(x) = 2$$

$$g'\left(f\right) = 4f^3$$

- The derivative of h is given as

$$h^{'(x)} = g^{'(f)} f^{'(x)} = \left(4f^3\right) \bullet 2 = 4(2x+1)^3 \bullet 2 = 8(2x+1)^3$$

Partial differentiation and the gradient

5.2 Partial Differentiation and Gradients

- Instead of considering $x \in \mathbb{R}$, we consider $\boldsymbol{x} \in \mathbb{R}^n$, e.g., $f(\boldsymbol{x}) = f(x_1,\ x_2)$
- For a function $f: \mathbb{R}^n \to \mathbb{R}, \boldsymbol{x} \mapsto f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n$ of n variables $x_1, \ldots, x_n$, we define the partial derivatives as

$$\frac{\partial f}{\partial x_1} := \lim_{h \to 0} \frac{f(x_1 + h, x_2, \ldots, x_n) - f(\boldsymbol{x})}{h} \quad = f_{x_1}$$

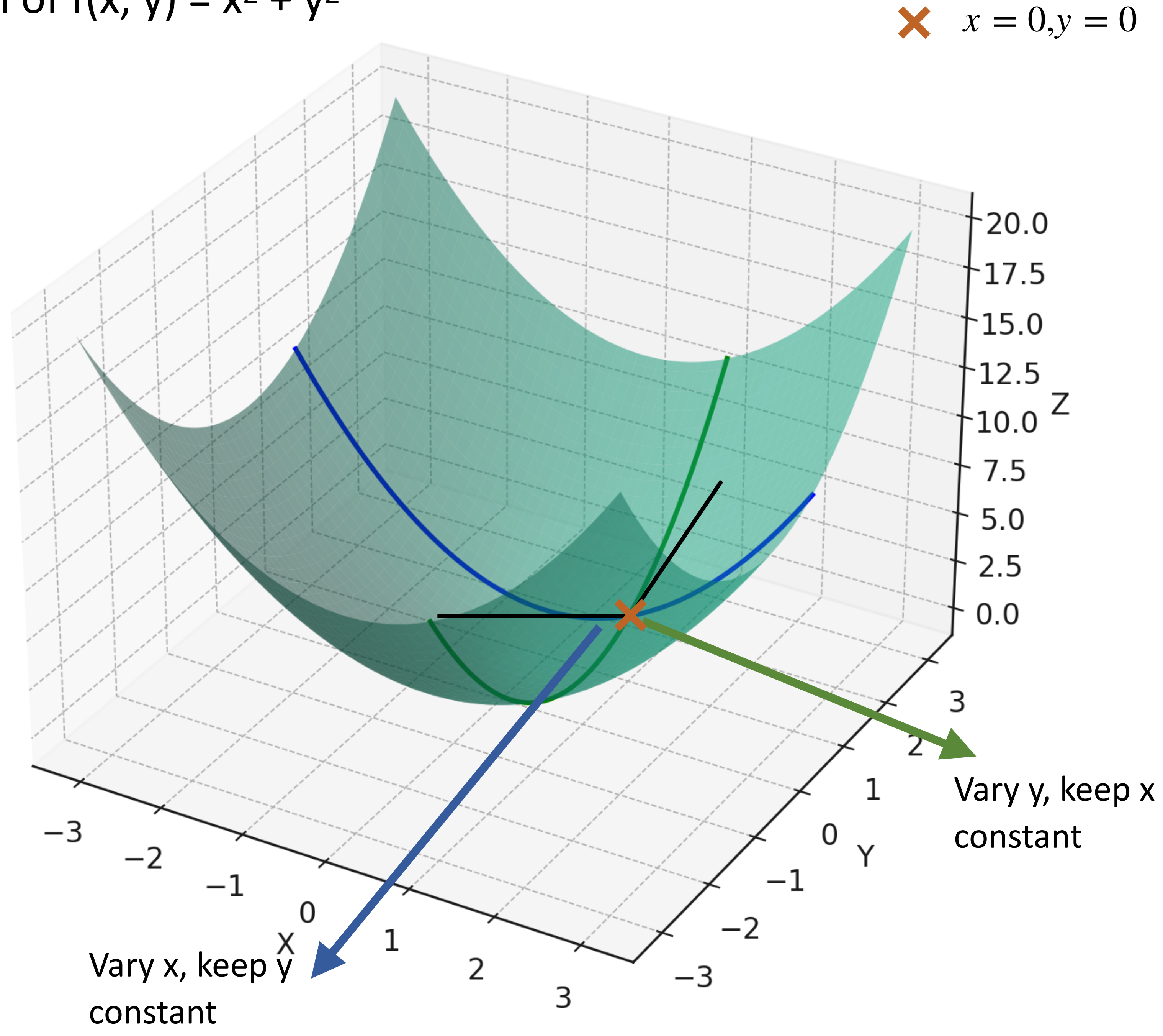
rate of change with respect to x_1

$$\vdots$$

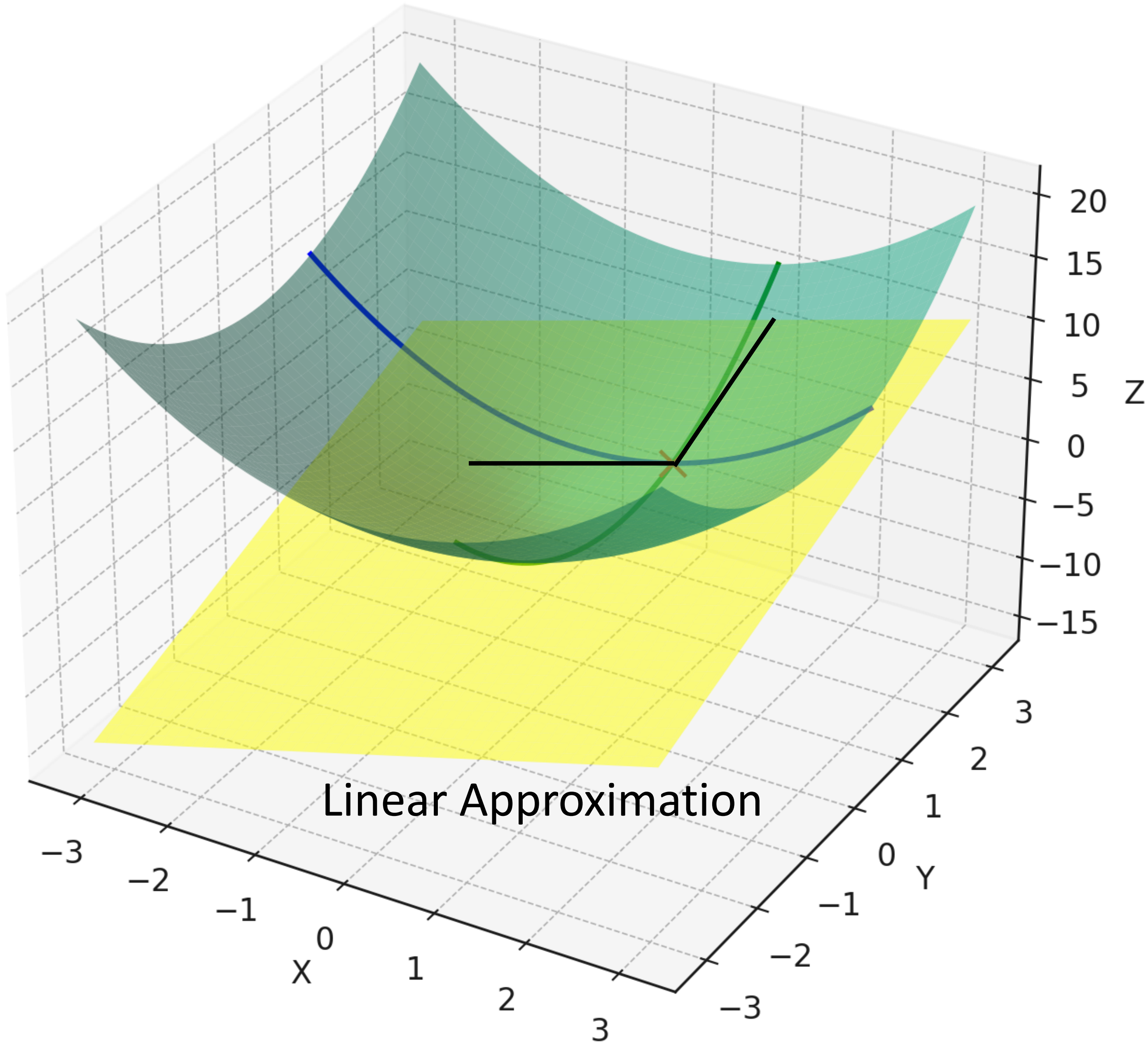
$$\frac{\partial f}{\partial x_n} := \lim_{h \to 0} \frac{f(x_1, \ldots, x_{n-1}, x_n + h) - f(\boldsymbol{x})}{h} \quad = f_{x_n}$$

rate of change with respect to x_n

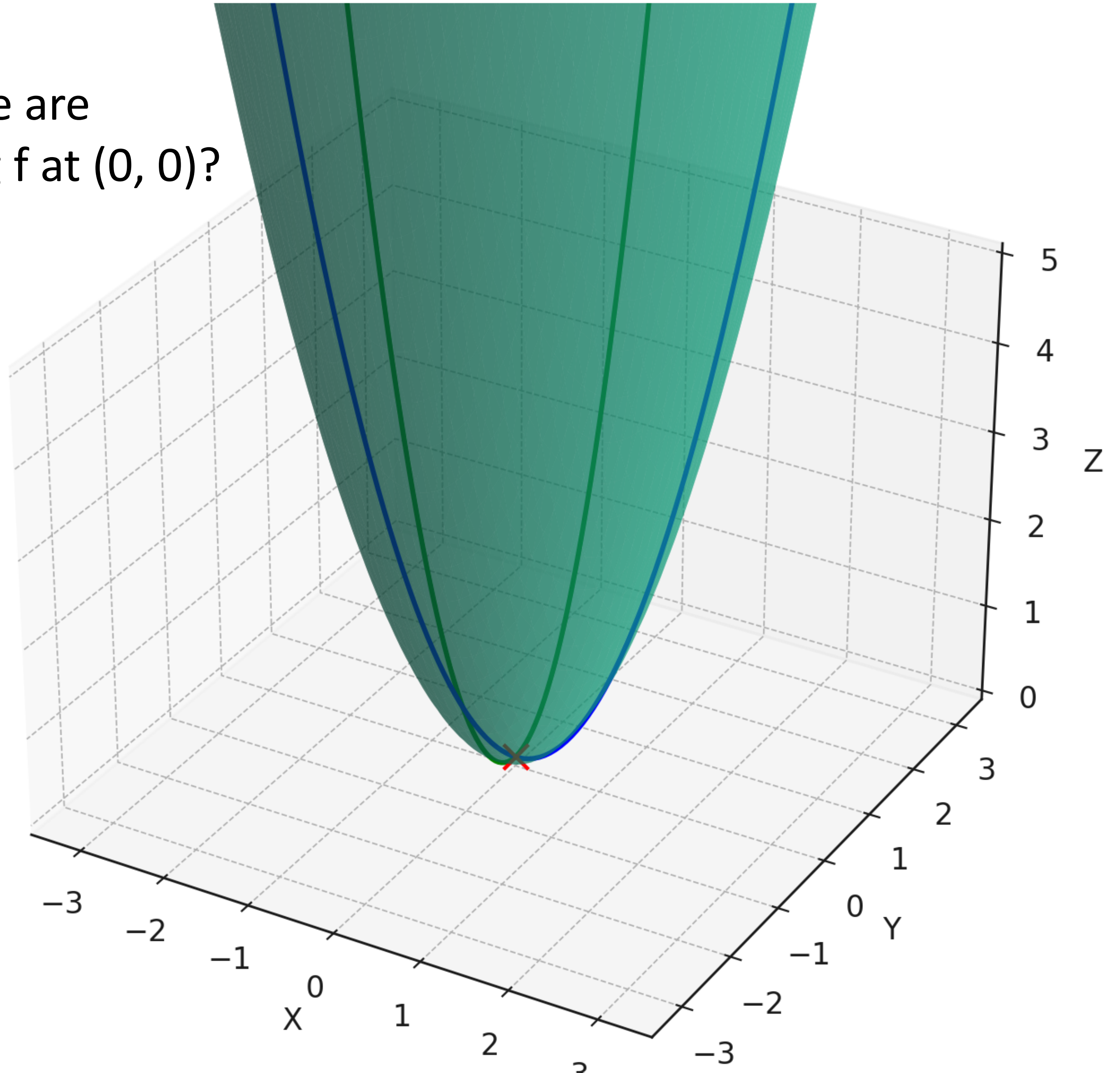
3D graph of $f(x, y) = x^2 + y^2$



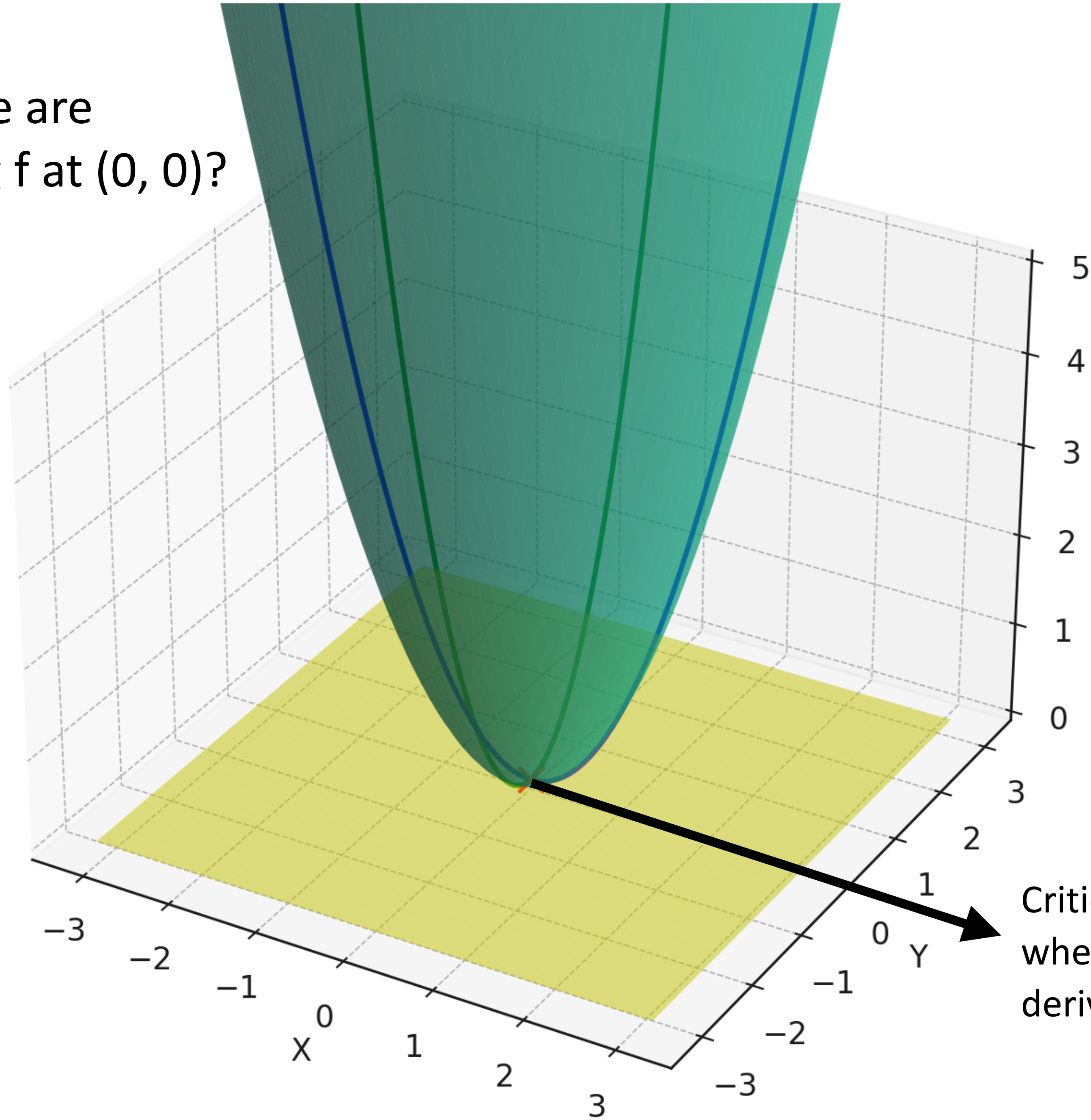
Tangent plane to the graph zoomed in



What if we are evaluating f at (0, 0)?



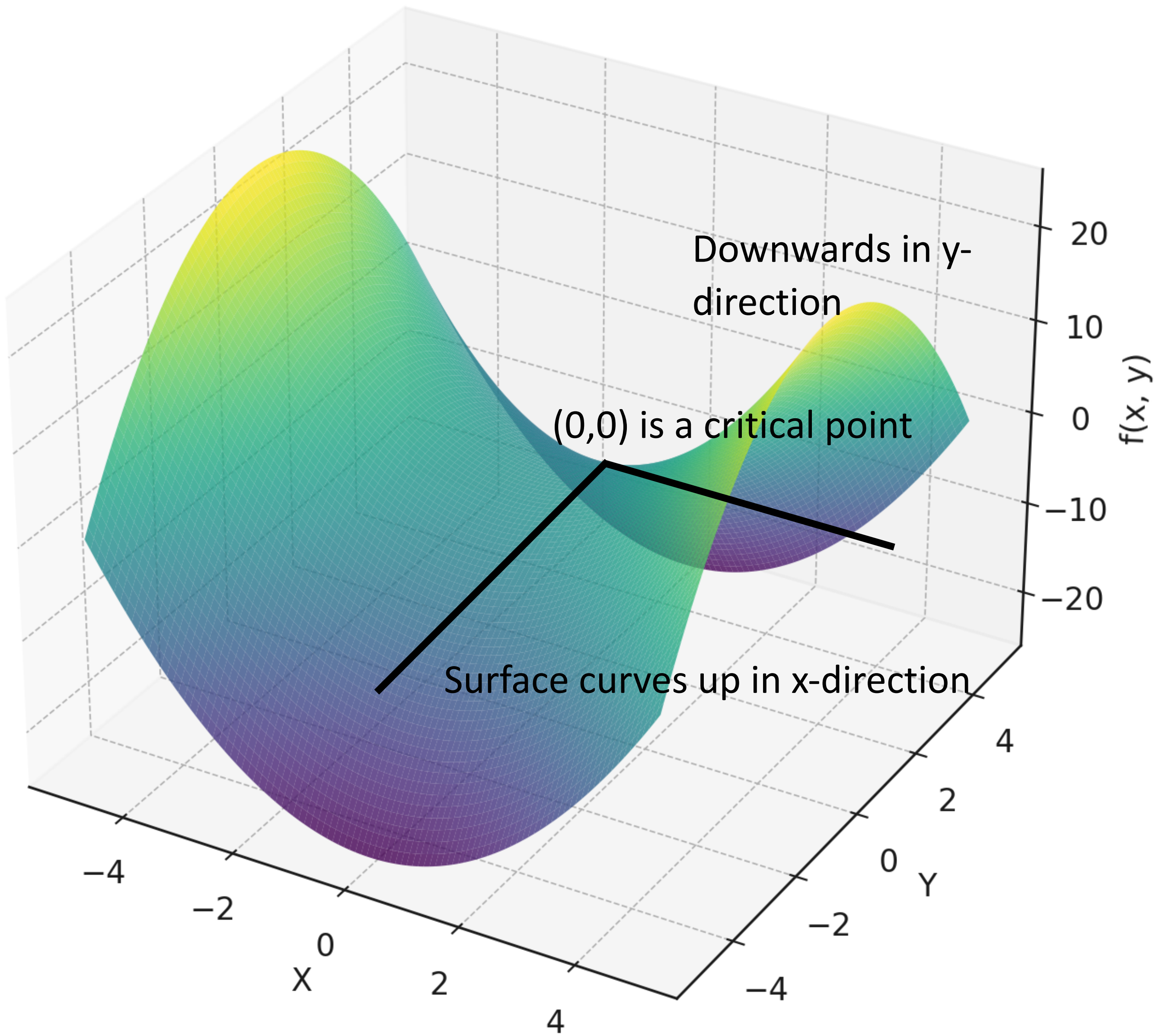
What if we are evaluating f at (0, 0)?



(0,0) is a global minimum

Critical point where all partial derivatives are 0

What about (0, 0) for $f(x, y) = x^2 - y^2$



(0,0) is a saddle point

5.2 Partial Differentiation and Gradients

- The generalization of the derivative to functions of several variables is the gradient.
- We find the gradient of the function f with respect to $\boldsymbol{x}$ by
  - varying one variable at a time and keeping the others constant.
  - The gradient is the collection of the partial derivatives.

- We collect the partial derivatives in the row vector

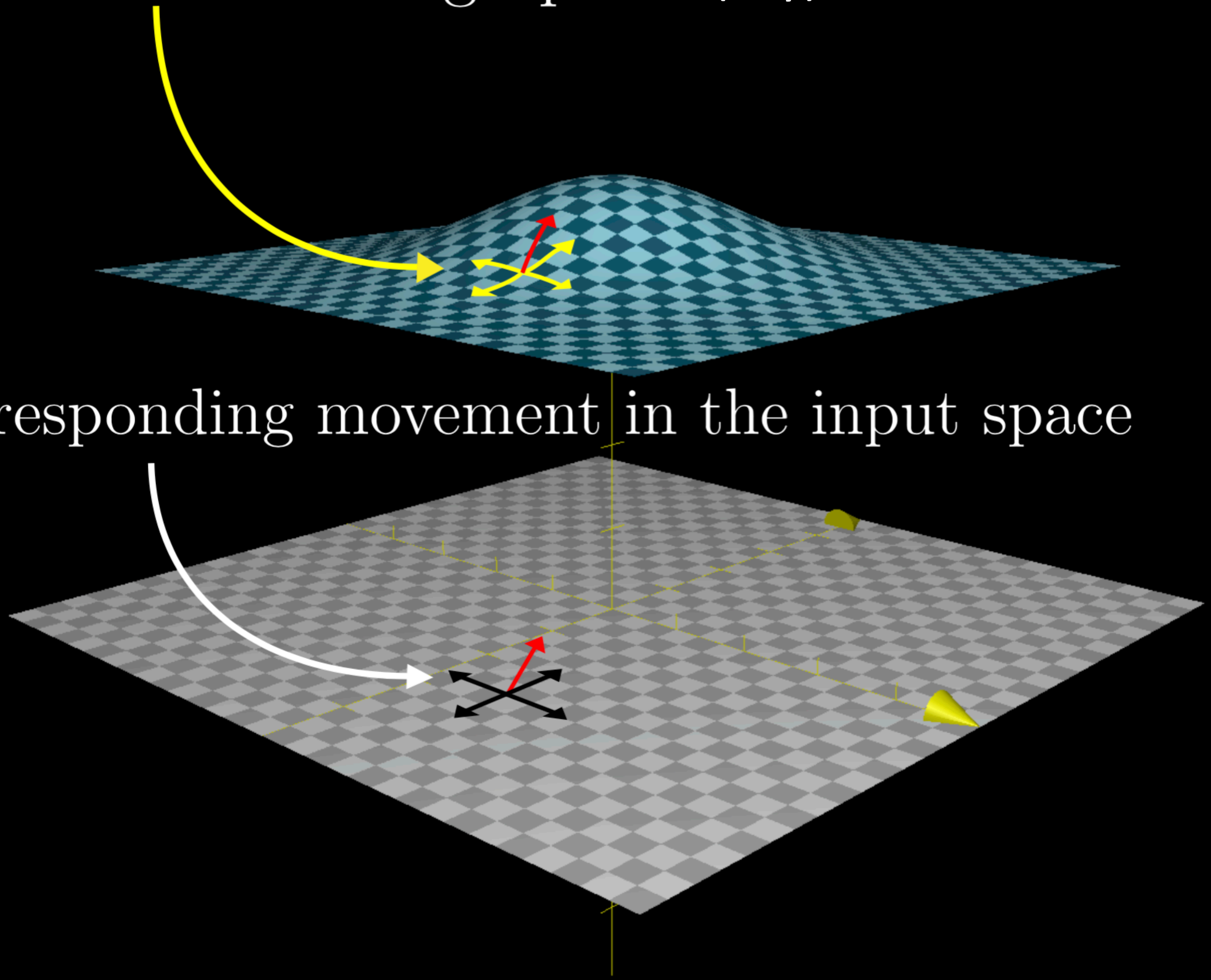
$$\nabla_x f = \operatorname{grad} f = \frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_n} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{1\times n}$$

5.2 Partial Differentiation and Gradients

- $$\nabla_x f = \operatorname{grad} f = \frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{1\times n}$$
- n is the number of variables and 1 is the dimension of the image/range/codomain of f.

- The row vector $\nabla_x f \in \mathbb{R}^{1\times n}$ is called the gradient of f or the Jacobian.

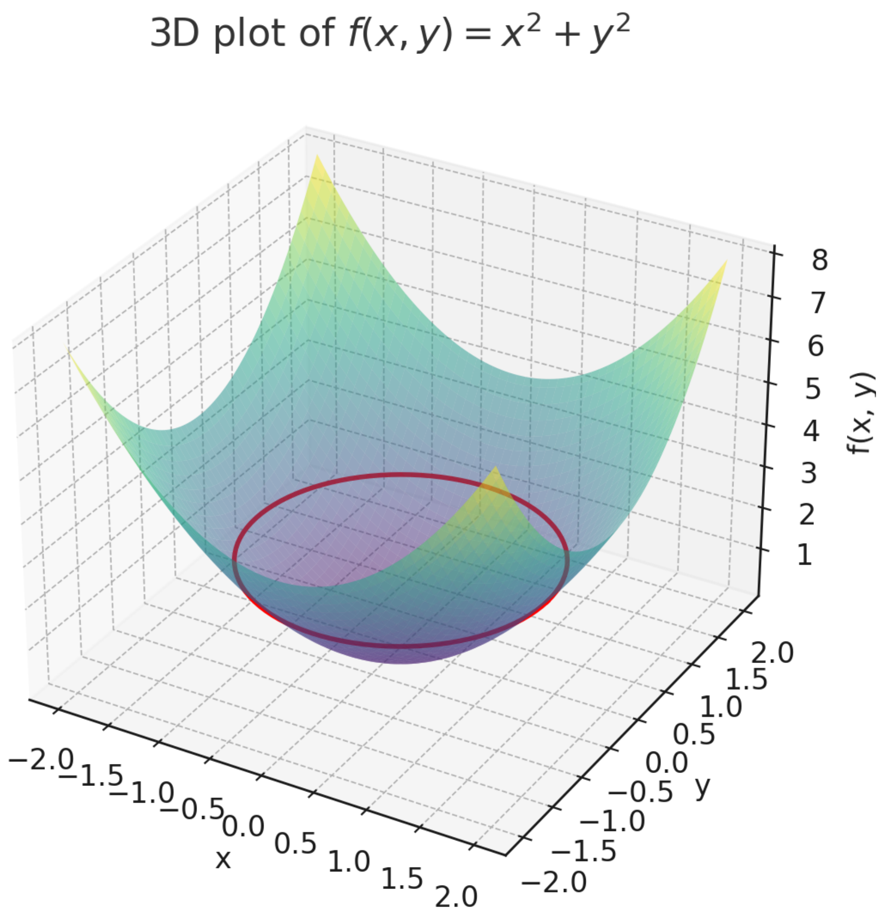
Movement on the graph of f(x, y)



Corresponding movement in the input space

Gradient points in the direction of steepest ascent

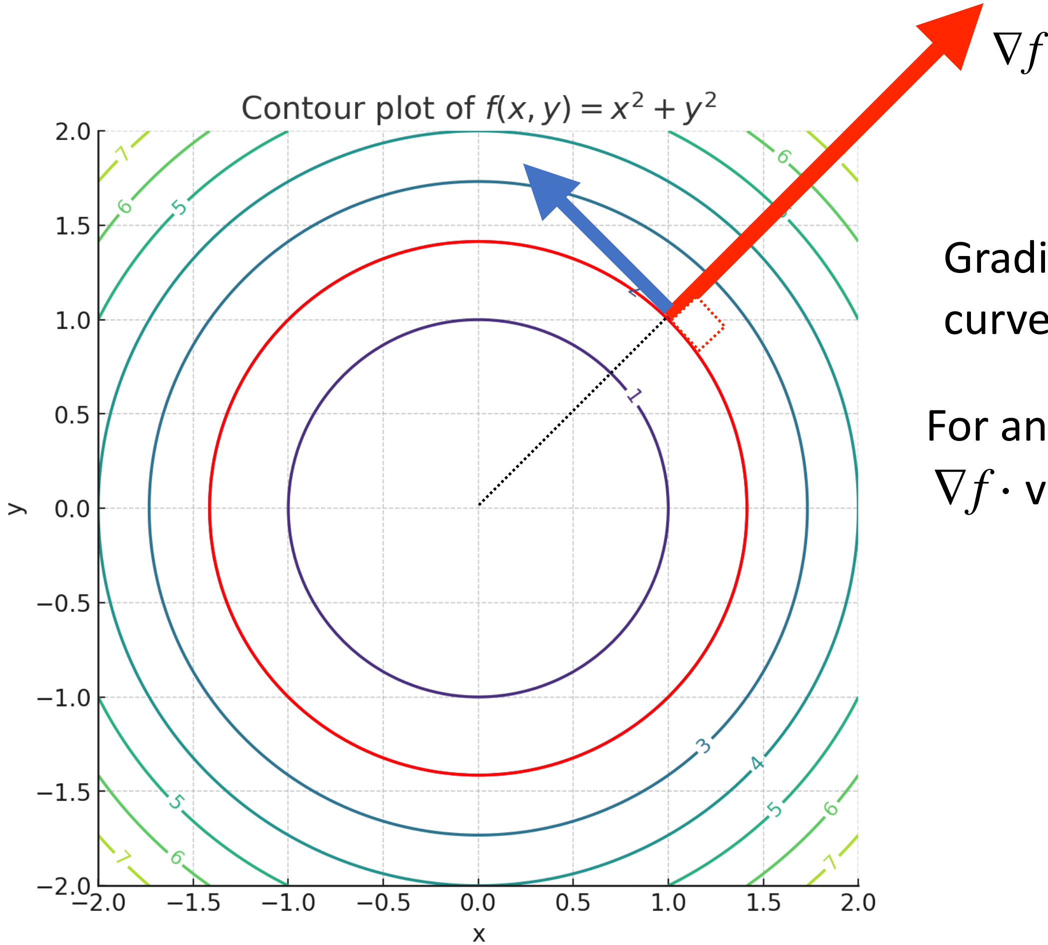
Let's go back to geometry



Define level curve (isocurve) =curve along which f has a constant value.

For our example, $x^2 + y^2 = 2$ defines a circle

What if we want to find the gradient of f? $\nabla f(x,y) = [\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}] = [2x \quad 2y]$



Gradient perpendicular to the level curve

For any vector tangent **v** to level,
$\nabla f \cdot \mathrm{v} = 0$

What about a function $f(x, y, z)$?

- Level surfaces defined by $f(x, y, z) = c$, where c is constant.

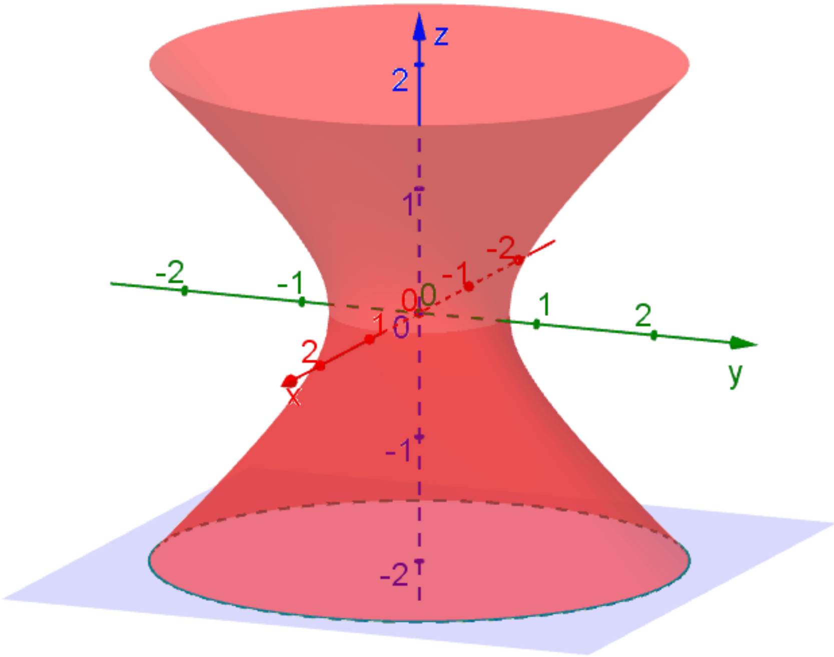
- For any vector v tangent to the level surface, the gradient is perpendicular to it:

$$\nabla f \cdot \mathrm{v} = 0$$

- The ∇f is perpendicular to the tangent plane to the level surface.
- The ∇f is normal vector to the tangent plane.

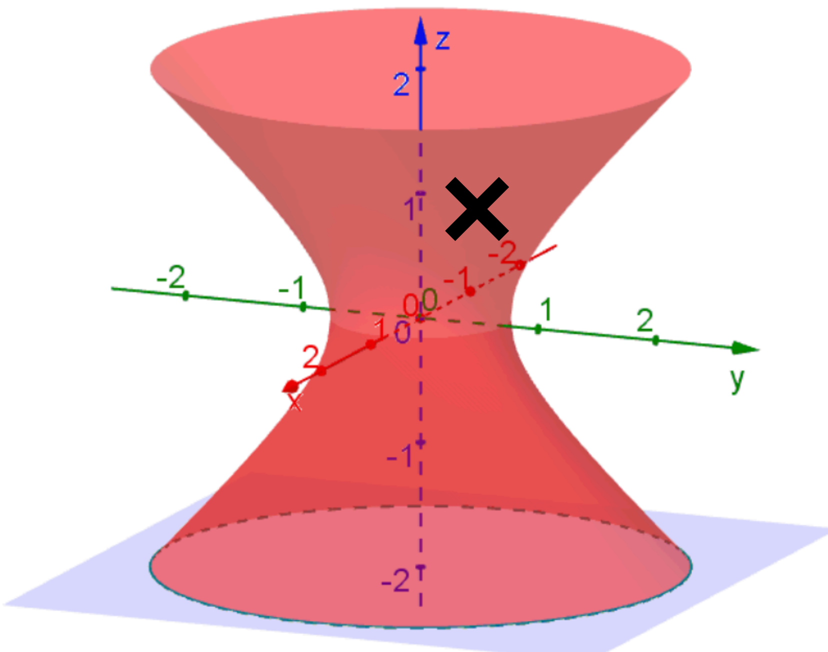
Let's take $f(x, y, z) = x^2 + y^2 - z^2$

- Let's choose a surface level $x^2 + y^2 - z^2 = 1$.



Let's take $f(x, y, z) = x^2 + y^2 - z^2$

- Let's choose a surface level $x^2 + y^2 - z^2 = 1$.



What is the tangent plane of this surface at some point (2, 1, 1)?

$$\nabla f = [2x \quad 2y \quad -2z]$$

$$\nabla f \cdot (\mathrm{w} - \mathrm{w}_0) = 0$$

$$4(x-2) + 2(y-1) - 2(z-1) = 0$$

5.2 Partial Differentiation and Gradients

- For $f(x_1, x_2) = x_1^2 x_2 + x_1 x_2^3 \in \mathbb{R}$, the partial derivatives (i.e., the derivatives of f with respect to x_1 and x_2 are

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 2x_1 x_2 + x_2^3$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = x_1^2 + 3x_1 x_2^2$$

- and the gradient is then

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \left[\frac{\partial f(x_1, x_2)}{\partial x_1} \quad \frac{\partial f(x_1, x_2)}{\partial x_2}\right] = \left[2x_1 x_2 + x_2^3 \quad x_1^2 + 3x_1 x_2^2\right] \in \mathbb{R}^{1\times 2}$$

5.2.1 Basic Rules of Partial Differentiation

- Product rule:

$$\frac{\partial}{\partial \boldsymbol{x}}\big(f(\boldsymbol{x})g(\boldsymbol{x})\big) = \frac{\partial f}{\partial \boldsymbol{x}}g(\boldsymbol{x}) + f(\boldsymbol{x})\frac{\partial g}{\partial \boldsymbol{x}}$$

- Sum rule:

$$\frac{\partial}{\partial \boldsymbol{x}}\big(f(\boldsymbol{x}) + g(\boldsymbol{x})\big) = \frac{\partial f}{\partial \boldsymbol{x}} + \frac{\partial g}{\partial \boldsymbol{x}}$$

- Chain rule:

$$\frac{\partial}{\partial \boldsymbol{x}}(g \circ f)(x) = \frac{\partial}{\partial \boldsymbol{x}}\big(g\big(f(\boldsymbol{x})\big)\big) = \frac{\partial g}{\partial f}\frac{\partial f}{\partial \boldsymbol{x}}$$

Tools in our maths toolbox

- Example - Partial Derivatives Using the Chain Rule
- For $f(x, y) = \left(x + 2y^3\right)^2$, we obtain the partial derivatives

$$\frac{\partial f(x, y)}{\partial x} = 2\left(x + 2y^3\right)\frac{\partial}{\partial x}\left(x + 2y^3\right) = 2\left(x + 2y^3\right)$$

$$\frac{\partial f(x, y)}{\partial y} = 2\left(x + 2y^3\right)\frac{\partial}{\partial y}\left(x + 2y^3\right) = 12\left(x + 2y^3\right)y^2$$

5.2.2 Chain Rule

- Consider a function $f: \mathbb{R}^2 \to \mathbb{R}$ of two variables x_1 and x_2.
- $x_1(t)$ and $x_2(t)$ are themselves functions of t.
- To compute the gradient of f with respect to t, we apply the chain rule:

$$\frac{df}{dt} = \frac{\partial f}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \frac{dx_1(t)}{dt} \\ \frac{dx_2(t)}{dt} \end{bmatrix} = \frac{\partial f}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dt}$$

Where df, dx denote a differential and ∂ refer to the partial derivatives.

$$\Delta f \approx f_{x_1} \Delta x_1 + f_{x_2} \Delta x_2$$

$$\frac{\Delta f}{\Delta t} \approx f_{x1} \frac{\Delta_{x_1}}{\Delta_t} + f_{x_2} \frac{\Delta_{x_2}}{\Delta_t}$$

when Δt goes to 0 : $$\frac{df}{dt} = f_{x_1} \frac{dx_1}{dt} + f_{x_2} \frac{dx_2}{dt}$$

Example

- Consider $f(x_1, x_2) = x_1^2 + 2x_2$, where $x_1 = \sin t$ and $x_2 = \cos t$, then

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t}$$
$$= 2\sin t\frac{\partial \sin t}{\partial t} + 2\frac{\partial \cos t}{\partial t}$$
$$= 2\sin t\cos t - 2\sin t = 2\sin t(\cos t - 1)$$

- The above is the corresponding derivative of f with respect to t.

5.2.2 Chain Rule

- If $f(x_1, x_2)$ is a function of x_1 and x_2, where $f: \mathbb{R}^2 \to \mathbb{R}$, $x_1(s,t)$ and $x_2(s,t)$ are themselves functions of two variables s and t, the chain rule yields the partial derivatives

$$\frac{\partial f}{\partial s} = \boxed{\frac{\partial f}{\partial x_1}} \boxed{\frac{\partial x_1}{\partial s}} + \boxed{\frac{\partial f}{\partial x_2}} \boxed{\frac{\partial x_2}{\partial s}}$$

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t}$$

- The gradient can be obtained by matrix multiplication

$$\frac{df}{d(s,t)} = \frac{\partial f}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{x}}{\partial (s,t)} = \underbrace{\begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}}_{= \frac{\partial f}{\partial \boldsymbol{x}}} \underbrace{\begin{bmatrix} \frac{\partial x_1}{\partial s} & \frac{\partial x_1}{\partial t} \\ \frac{\partial x_2}{\partial s} & \frac{\partial x_2}{\partial t} \end{bmatrix}}_{= \frac{\partial \boldsymbol{x}}{\partial (s,t)}}$$