Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- Univariate Calculus
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- **Multivariate Calculus**
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - **Linear approximations and Alternate gradient interpretations**
  - Applications/Advanced rules

Gradients and Linear Approximations

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

$$f(x) \approx \underbrace{f(x^*) + f'(x^*)(x - x^*)}_{L_{x^*}[f](x)}$$

around $x = x^*$

$$v \in \mathbb{R}^d, \quad x \in \mathbb{R}^d$$

$$f(x) \approx f(v) + \nabla f(v)^T (x - v)$$

$$= \underbrace{f(v) + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(v) \cdot (x_i - v_i)}_{L_v[f](x)}$$

around $x = v$

Gradients and Linear Approximations

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

$$f(y_1, v_2) \approx f(v_1, v_2) + \frac{\partial f}{\partial x_1}(v) \cdot (y_1 - v_1)$$

$$f(y_1, v_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_1}(v)(y_1 - v_1)$$

$$f(v_1, y_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_2}(v)(y_2 - v_2)$$

$$f(y_1, y_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_1}(v)(y_1 - v_1) + \frac{\partial f}{\partial x_2}(v)(y_2 - v_2)$$

$$f(y_1, y_2) \approx f(v_1, v_2) + \nabla f(v)^T (y - v)$$

Gradients and Linear Approximations

$$f(x_1, x_2) = x_1^2 + x_2^2 \qquad \nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

(i) Approximate f around $(6, 2)$

$$f(v) = 40, \quad \nabla f(v) = \begin{bmatrix} 12 \\ 4 \end{bmatrix}$$

$$f(x) \approx 40 + [12, 4] \begin{bmatrix} x_1 - 6 \\ x_2 - 2 \end{bmatrix}$$

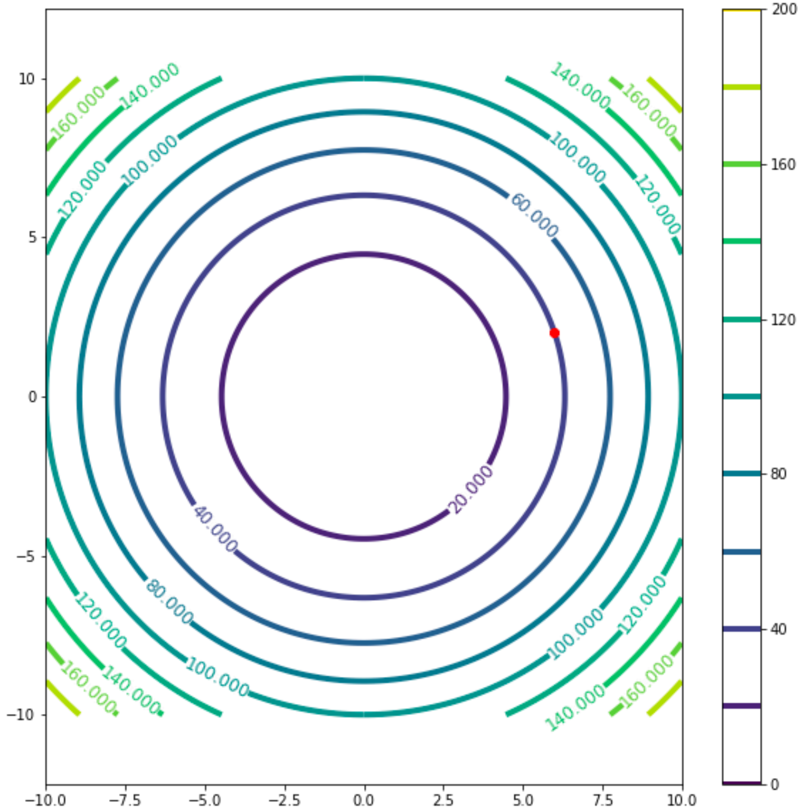
$$= 40 + 12(x_1 - 6) + 4(x_2 - 2)$$

$$= 40 + 12x_1 + 4x_2 - 72 - 8$$

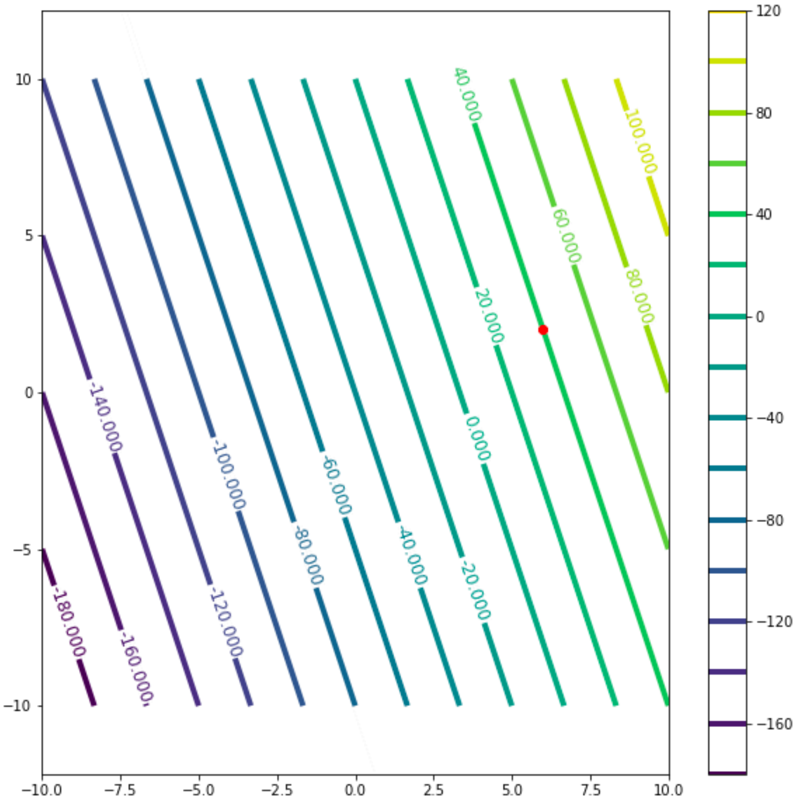
$$= 12x_1 + 4x_2 - 40$$

$(x_1, x_2) \approx (6, 2)$

Gradients and Linear Approximations

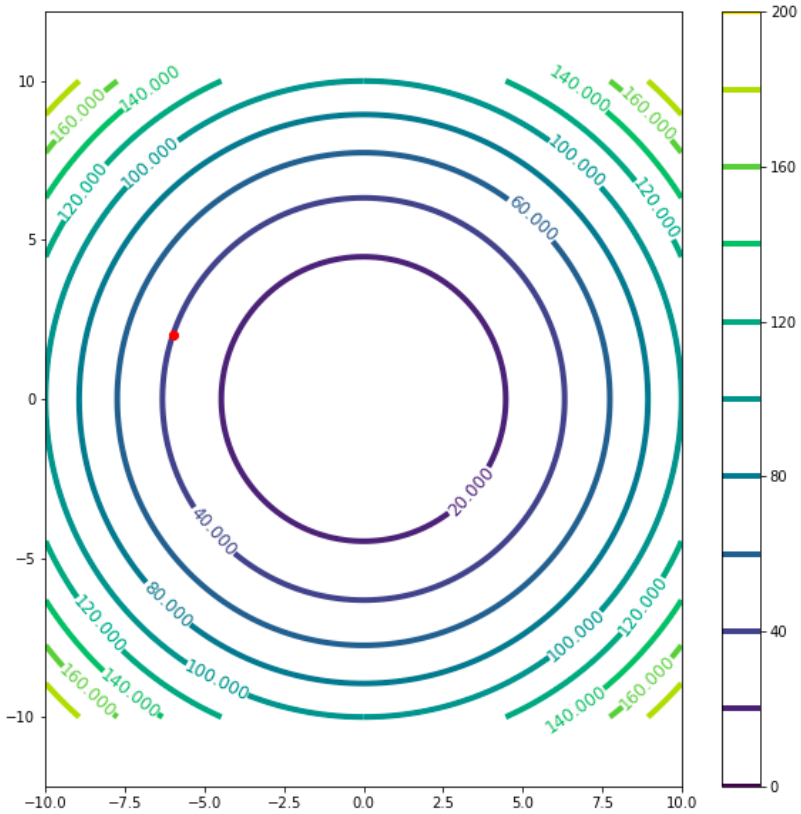


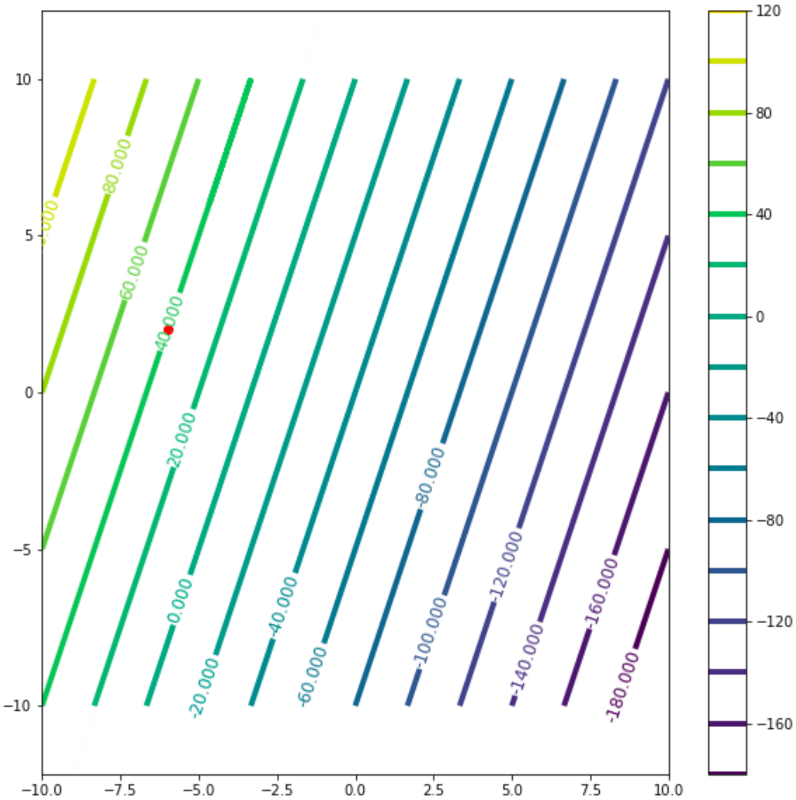
$f(x) = x_1^2 + x_2^2$



$L_v[f]$

Gradients and Linear Approximations



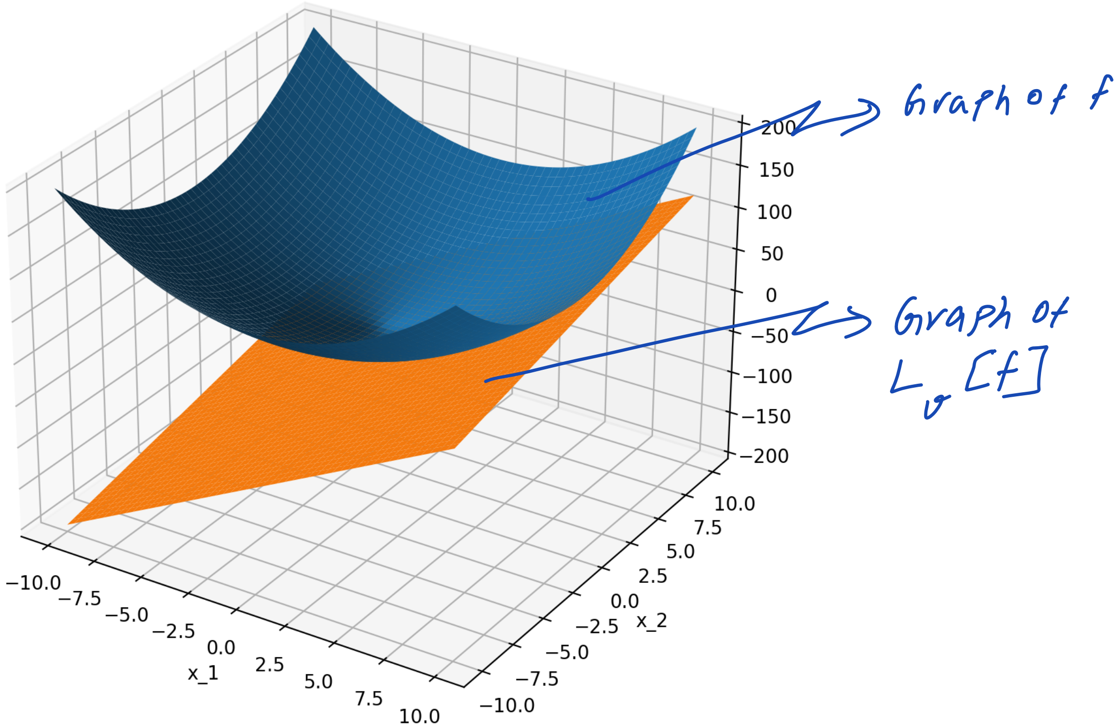


$f(x) \approx L_v[f](x) \quad (x \approx v)$

Gradients and Tangent Planes

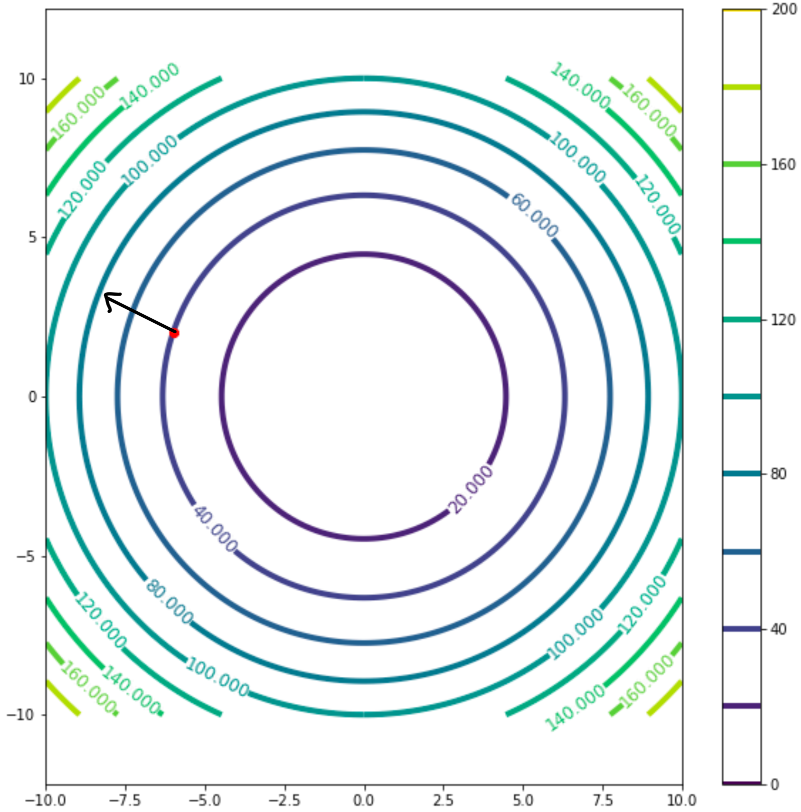
$f(x) = x_1^2 + x_2^2$

The graph of $L_v[f]$ is a plane that is tangent to the graph of f at the point $(v, f(v))$



Gradients and Contours

$$v = \begin{bmatrix} -6 \\ 2 \end{bmatrix} \quad \nabla f(v) = \begin{bmatrix} -12 \\ 4 \end{bmatrix}$$



$$\nabla f(v) \perp \{x \in \mathbb{R}^d : f(x) = f(v)\}$$

$$\nabla f(v) \perp \{x \in \mathbb{R}^d : L_v[f](x) = f(v)\}$$

$$\|$$

$$\{x \in \mathbb{R}^d : f(v) + \nabla f(v)^T (x - v) = f(v)\}$$

$$\|$$

$$\{x \in \mathbb{R}^d : \nabla f(v)^T x = \nabla f(v)^T v\}$$

$$\{x \in \mathbb{R}^d : w^T x = b\}$$

Directional Derivative

$$D_u[f](v) = \lim_{\alpha \to 0} \frac{f(v+\alpha u) - f(v)}{\alpha}$$

Directional derivative of f at the point v, along u.

$$= \lim_{\alpha \to 0} \frac{f(v) + \nabla f(v)^T \alpha u - f(v)}{\alpha}$$

$$= \nabla f(v)^T u$$

Cauchy-Schwarz Inequality

$a_1, a_2 \cdots a_d$

$b_1, b_2 \cdots b_d$

$\|a\| = \sqrt{a_1^2 + \cdots + a_d^2}$

$$-\|a\| \cdot \|b\| \le a^T b \le \|a\| \|b\|$$

$a = \alpha b$

$\alpha < 0$

$a = \alpha b$

$\alpha > 0$

Direction of Steepest Ascent

f

Find a direction u, that maximises the rate of change of f as you move from v along u.

Maximise $D_u[f](v)$

Find $u \in \mathbb{R}^d$, $\|u\| = 1$ and which maximises

$$D_u[f](v)$$

$$= \nabla f(v)^T u$$

$$u = \alpha \cdot \nabla f(v)$$

Descent Directions

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

$$v \in \mathbb{R}^d .$$

What are the valid directions, such that f decreases

For what values of u : $D_u[f](v) < 0$

$$\Updownarrow$$

$$\nabla f(v)^T u < 0$$

Descent directions : $\{ u \in \mathbb{R}^d : \nabla f(v)^T u < 0 \}$

Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- Univariate Calculus
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- **Multivariate Calculus**
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - Linear approximations and Alternate gradient interpretations
  - **Applications/Advanced rules**

Higher Order Approximations

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

$$f(x) \approx f(v) + \nabla f(v)^T (x - v)$$

(Valid around $x = v$)

$$f(x) \approx f(v) + \nabla f(v)^T (x - v) + \frac{1}{2} (x - v)^T \underbrace{\nabla^2 f(v)}_{\substack{d \times d \text{ matrix} \\ \text{Hessian}}} (x - v)$$

Higher Order Approximations

Maxima, minima and saddle points

If $f(x)$ is minimised
at v

$\Downarrow$

$\nabla f(v) = 0$

$\{v : \nabla f(v) = 0\} \rightarrow$ Critical point