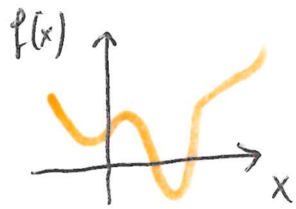
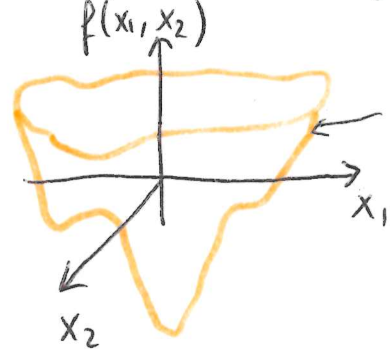
ESE 415 — Lecture 3 — 9/4/2024

Goal: Extend last week's results for $f : \mathbb{R} \to \mathbb{R}$ to the multivariate case $f : \mathbb{R}^n \to \mathbb{R}$





2d surface in 3d space

Summary of some prerequisties: Let $f : \mathbb{R}^n \to \mathbb{R}$ be sufficiently smooth (higher order derivatives all exist and are continuous).

- Directional derivative of f along a direction $v \in \mathbb{R}^n$ $(\|v\| = 1)$

$$(D_v f)(x) := \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t}$$

- Partial derivatives of f: Special cases of directional derivatives when $v = \underbrace{\begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix}}_{e_1}, \underbrace{\begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix}}_{e_2}, \dots, \underbrace{\begin{pmatrix}0\\\vdots\\0\\1\end{pmatrix}}_{e_n}$

$$\underbrace{\left(\frac{\partial}{\partial x_i} f\right)(x)}_{\frac{\partial f}{\partial x_i}} := (D_{e_i} f)(x)$$

- <u>Gradient</u> of $f: \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ ("total derivative")

$$(\nabla f)(x) = \begin{pmatrix} \left(\frac{\partial f}{\partial x_1}\right)(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

Recall that if gradient exists, then $(D_v f)(x) = (\nabla f)(x)^T v$.
(sufficient smoothness)

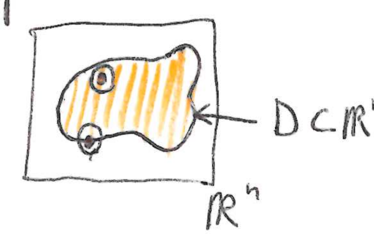
- <u>Hessian</u> of $f: \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ (second derivative):

$$(Hf)(x) = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} f, & \frac{\partial^2}{\partial x_1 \partial x_2} f, & \dots, & \frac{\partial^2}{\partial x_1 \partial x_n} f \\ \frac{\partial^2}{\partial x_1 \partial x_2} f, & \dots & & \\ \vdots & & & \\ \frac{\partial^2}{\partial x_1 \partial x_n} f, & \dots & & \frac{\partial^2}{\partial x_n^2} f \end{pmatrix}$$

Recall that under sufficient smoothness assumptions, the Hessian matrix is always a symmetric matrix (Schwarz's theorem)

- <u>Misc.</u>: • $B_\varepsilon(x) := \{ y \in \mathbb{R}^n \mid \|y - x\| < \varepsilon \}$

  • $x \in \mathbb{R}^n$ is an interior point of a set $D \subset \mathbb{R}^n$
  $\overset{\text{def.}}{\iff} \exists_{\varepsilon > 0}\ B_\varepsilon(x) \subset D.$



Definition (Minimizers): Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$.

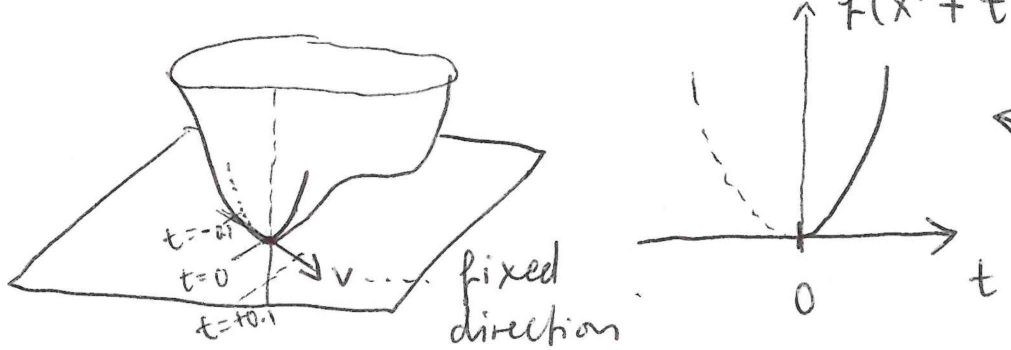
A point $x^* \in D$ is called

1. global minimizer of f on D if $f(x^*) \leq f(x)$ $\forall x \in D$.
2. strict global min. of f on D if $f(x^*) < f(x)$ $\forall x \in D \setminus \{x^*\}$
3. local min. of f if $\exists \delta > 0$ $\forall x \in \underbrace{B_\delta(x^*) \cap D}_{\subseteq D}$ $f(x^*) \leq f(x)$.
4. strict local min. of f if $\exists \delta > 0$ $\forall x \in (B_\delta(x^*) \cap D) \setminus \{x^*\}$ $f(x^*) < f(x)$.
5. stationary point of f if $(\nabla f)(x^*)$ exists and $(\nabla f)(x^*) =$

Theorem (First-order necessary condition)

Let $f \in C^1(D, \mathbb{R}^n)$. If x^* is a local minimizer of f and also an interior point of D, then $(\nabla f)(x^*) = 0$.

Proof: The idea is to link the new multivariate case to the univariate case that we now understand.



← back in the one-dim. case !!

$$\rightsquigarrow \underbrace{(D_v f)(x^*)}_{(\nabla f)(x^*)^T v} = 0 \quad \forall v \in \mathbb{R}^n \Rightarrow (\nabla f)(x^*) = 0.$$