117

so

$$J_{f_1 f_2}(x) = J_m(f(x)) \cdot J_f(x) \qquad \text{[chain rule]}$$

$$= \underbrace{\begin{pmatrix} v, & u \end{pmatrix}}_{\text{at } f(x)} \cdot \underbrace{\begin{pmatrix} \partial_1 f_1 & \cdots\cdots & \partial_n f_1 \\ \partial_1 f_2 & \cdots\cdots & \partial_n f_2 \end{pmatrix}}_{\text{at } x}$$

$$= \big(f_2(x), f_1(x)\big) \cdot \begin{pmatrix} \partial_1 f_1 & \cdots\cdots & \partial_n f_1 \\ \partial_1 f_2 & \cdots\cdots & \partial_n f_2 \end{pmatrix}$$

$$= \big(f_2 \partial_1 f_1 + f_1 \partial_1 f_2, \ \cdots\cdots, \ f_2 \partial_n f_1 + f_1 \partial_n f_2\big)$$

$$\Longleftrightarrow \quad \frac{\partial(f_1 f_2)}{\partial x_1} = f_2 \frac{\partial f_1}{\partial x_1} + f_1 \frac{\partial f_2}{\partial x_1} \qquad \text{(Leibniz rule)}$$

(3)

$$\begin{tikzcd} I \arrow[r, "f"] \arrow[rr, bend left, "g\circ f"] & \mathbb{R}^m \arrow[r, "g"] & \mathbb{R} \end{tikzcd}$$

$I \subset \mathbb{R}$

$\llcorner$ variables $y_1, \dots, y_m$

$$f(t) = (f_1(t), \dots, f_m(t))$$

$$g\circ f : I \longrightarrow \mathbb{R} \quad \left(\begin{array}{l} \text{one variable} \\ \text{one value} \end{array} \right.$$

If f and g are differentiable [f has a derivative $f'(t) = (f_1'(t), \dots, f_m'(t))$]

then $g\circ f$ has a derivative and

$$(g\circ f)'(t) = \frac{\partial g}{\partial y_1}(f(t))\, f_1'(t) + \dots + \frac{\partial g}{\partial y_m}(f(t))\, f_m'(t)$$

This is also a case of the chain rule: (119)

$$J_f(t) = \begin{pmatrix} f_1' \\ \vdots \\ f_m' \end{pmatrix}, \quad J_g\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \left(\frac{\partial g}{\partial y_1} \cdots \frac{\partial g}{\partial y_m}\right)$$

so

$$J_g(f(t)) \cdot J_f(t) = f_1' \frac{\partial g}{\partial y_1} + \cdots + f_m' \frac{\partial g}{\partial y_m}$$

evaluated at t

evaluated at $f(t)$

$$J_{g\circ f}(t) = ((g\circ f)'(t)) = J_g(f(t)) \cdot J_f(t)$$

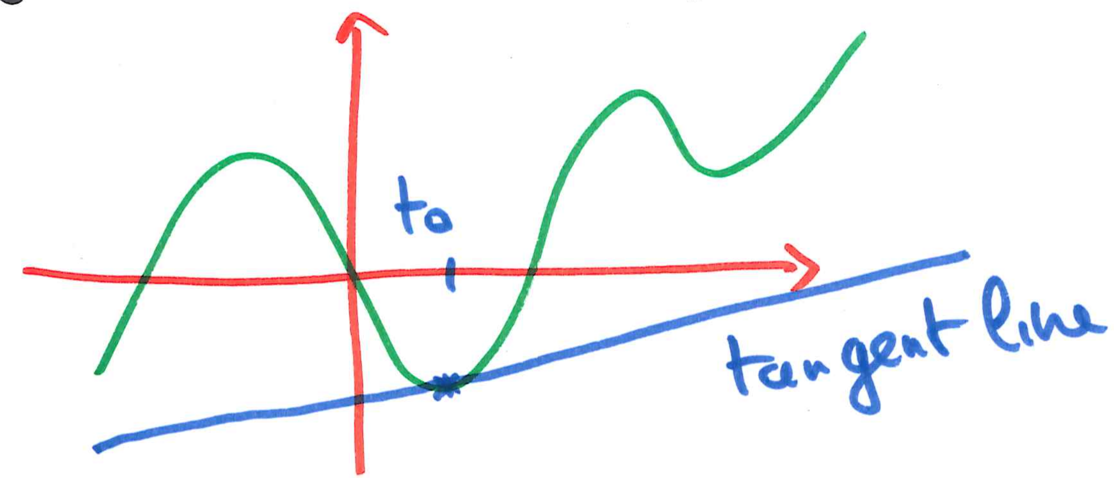
$$\boxed{(g\circ f)'(t) = \nabla g(f(t)) \cdot f'(t)}$$

scalar product

Some geometric interpretations or properties of the differential

<u>Generalize:</u> . the derivative $f'(t_0)$ of $f : I \longrightarrow \mathbb{R}$, $t_0 \in I$, gives the equation of the tangent line at $(t_0, f(t_0))$ to the graph of f:

$$y - f(t_0) = f'(t_0)(x - t_0)$$



- If $f'(t_0) > 0$ then f is strictly increasing "at t_0"

Def. $X \subset \mathbb{R}^n$ open

$f : X \longrightarrow \mathbb{R}^m$

Graph of f : $\Gamma_f = \{(x, f(x)) \mid x \in X\} \subset \mathbb{R}^{n+m}$

If f is differentiable at x_0 then the graph of the affine-linear map

$$g(x) = f(x_0) + J_f(x_0) \cdot (x - x_0)$$

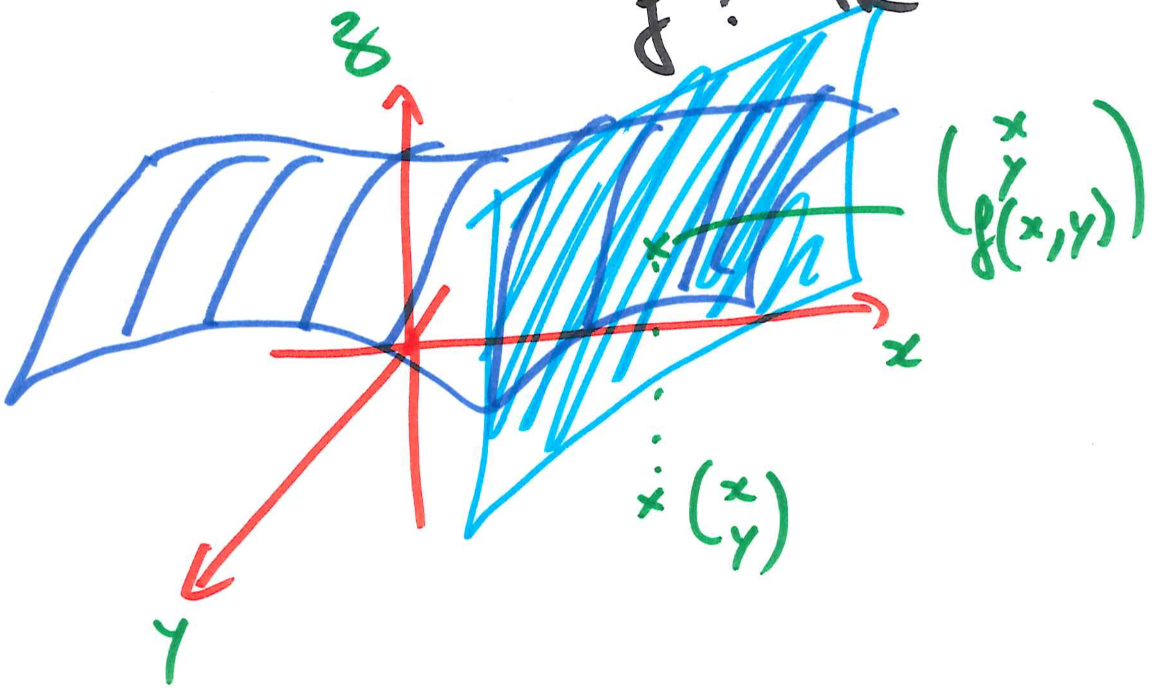
matrix product

is an affine subspace of $\mathbb{R}^{n+m}$ of dimension n called the <u>tangent space</u> to the graph of f at $(x_0, f(x_0))$.

$g: \mathbb{R}^n \longrightarrow \mathbb{R}^m$, graph $= \{(x, g(x)) \mid x \in \mathbb{R}^n\}$

<u>Ex.</u> $n = 2$, $m = 1$

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



$\Gamma_f =$ " surface " in $\mathbb{R}^3$

Case $m=1$:

$$f : X \longrightarrow \mathbb{R}, \quad X \subset \mathbb{R}^n$$

The ~~equation~~ tangent space to the graph of f at x_0 is the set of vectors $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^{n+1}$

$x \in \mathbb{R}^n$, $y \in \mathbb{R}$

s.t.

$$y = f(x_0) + J_f(x_0) \cdot (x - x_0)$$

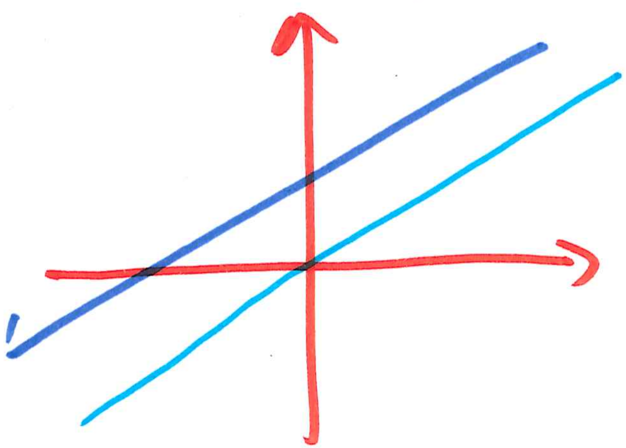
matrix product

$$= f(x_0) + Df(x_0) \cdot (x - x_0)$$

scalar product

$$y = f(x_0) + \nabla f(x_0) \cdot (x - x_0)$$

This is an affine - linear space of dimension n in $\mathbb{R}^{n+1}$. The linear space that is "parallel" to it is the set of $\binom{x}{y} \in \mathbb{R}^{n+1}$ s.t.



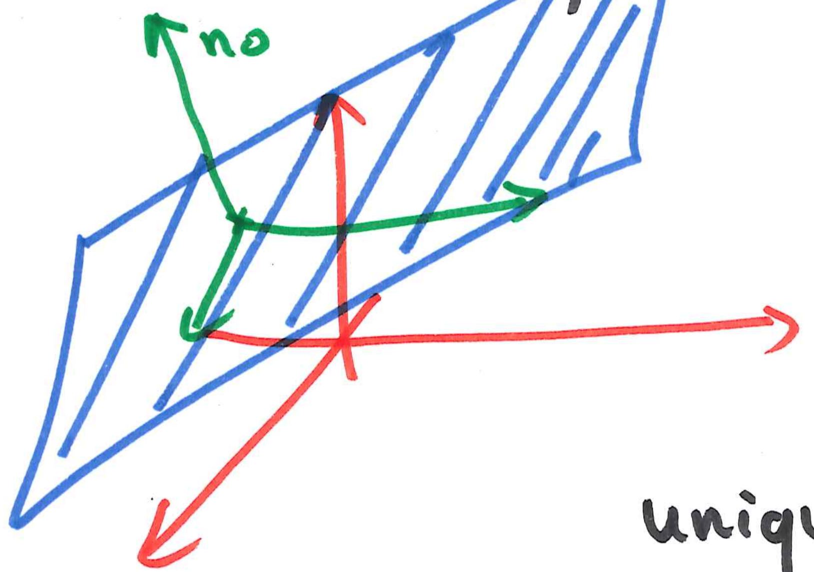
$$y = \nabla f(x_0) \cdot \text{[redacted]} \; x$$

<u>Geometric interpretation:</u>

This linear space is the space of $\binom{x}{y} \in \mathbb{R}^{n+1}$ <u>orthogonal</u> to

$$n_0 = (-\nabla f(x_0), 1) = \begin{pmatrix} -\nabla f(x_0) \\ 1 \end{pmatrix}$$

[because $y - \nabla f(x_0) \cdot x = \binom{x}{y} \cdot n_0$]

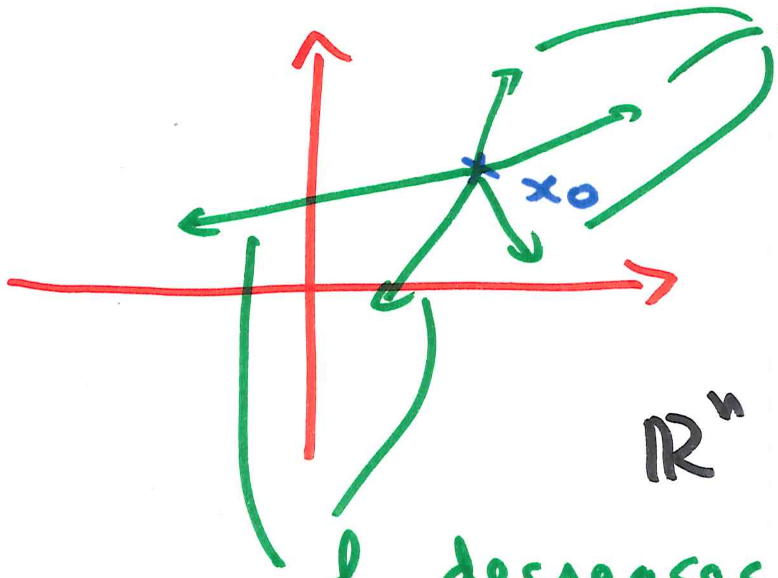


<u>Advantage of this description:</u>

The vector n_0 is unique (up to multiplication by $\neq 0$ real number)

Another interpretation of the gradient

$$f : \begin{array}{c} X \\ \cap \\ \mathbb{R}^n \end{array} \longrightarrow \mathbb{R}$$

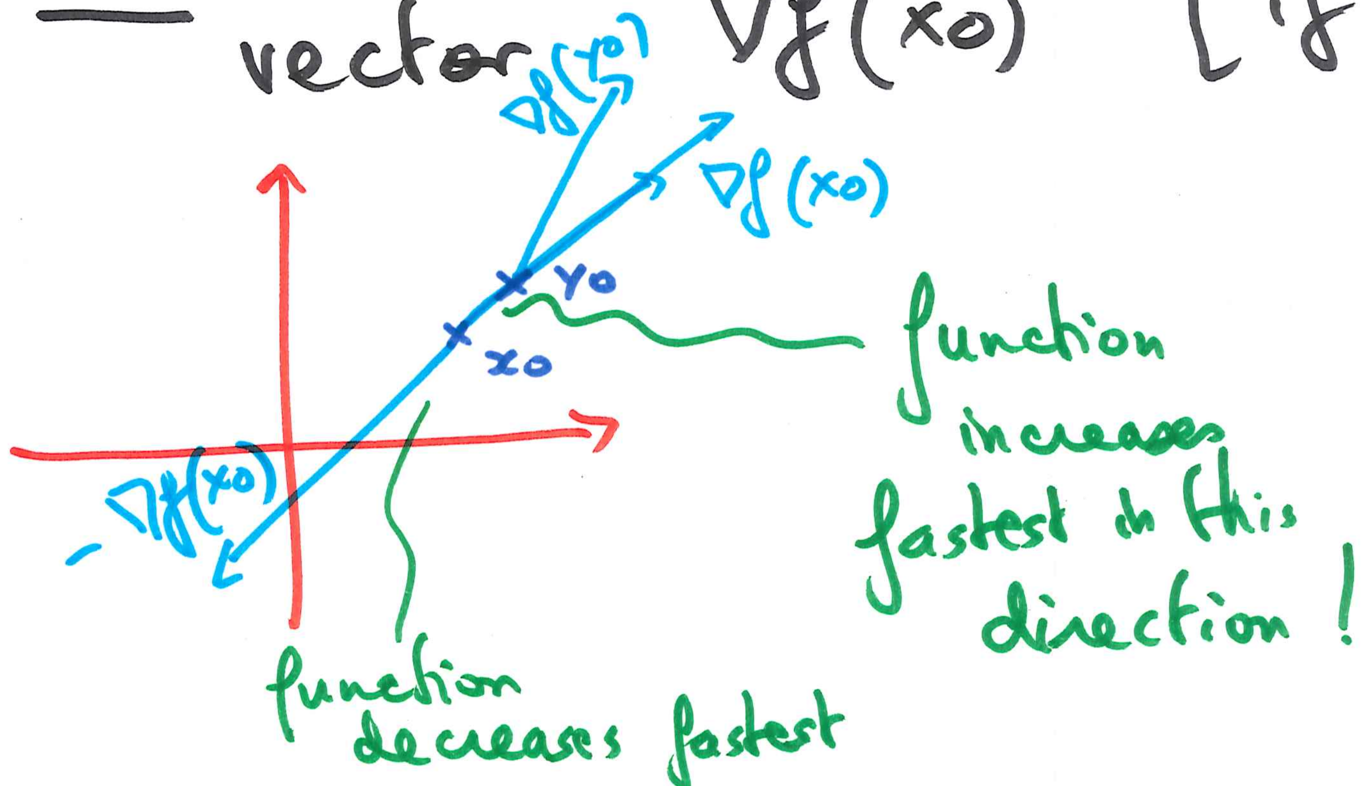


f increases (directional derivative >0)

f decreases (directional derivative <0)

Q. In which direction is the function "increasing fastest" around x_0?

A. It is given by the gradient vector $\nabla f(x_0)$ [if it is non-zero]

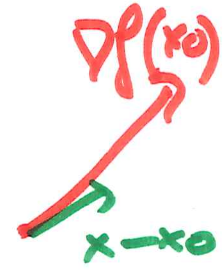


Reason:

$$f(x) - f(x_0) \approx \nabla f(x_0) \cdot (x - x_0)$$

scalar product

Want maximize as $x - x_0$ varies in all directions length 1



Cauchy - Schwarz inequality:

$$|\nabla f(x_0) \cdot (x - x_0)| \leq \|\nabla f(x_0)\| \cdot 1$$

with equality $\Leftrightarrow$ $x - x_0$ is proportional to $\nabla f(x_0)$

This gives >0 ~~rate~~ maximal rate of increase if $x - x_0 = \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|}$

(provided $\nabla f(x_0) \neq 0$!)

3.5. Higher derivatives

Recall: $$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$$

and so on, with $\frac{\partial}{\partial x_i^2}$ abbreviating $\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_i}\right)$

Def. $X \subset \mathbb{R}^n$ open

$f : X \longrightarrow \mathbb{R}^m$, $f = (f_1, \ldots, f_m)$

(1) (f is of <u>class C^1</u>) if (f is differentiable on X <u>and</u>

$J_f : X \longrightarrow \underset{\substack{n \text{ columns} \\ m \text{ rows}}}{\mathrm{Mat}(\mathbb{R})}$

is continuous) $\left[\boxed{\Longleftrightarrow \ \frac{\partial f_i}{\partial x_j} \text{ are continuous}} \right]$

(2) f is of class C^k, $k \geq 2$, if all partial derivatives of order 1 exist and all

are of class C^{k-1}.

<u>Notation:</u> $C^k(X, \mathbb{R}^n)$

$f \in C^\infty(X, \mathbb{R}^n) \Leftrightarrow$ in C^k for all $k \geq 1$

<u>Fact:</u> All simple elementary concrete functions are C^∞.

<u>Remark.</u> $f \in C^2(X, \mathbb{R}^n) \Rightarrow \frac{\partial f_i}{\partial x_j}$ are of class $C^1 \Rightarrow$ are differentiable

$\Rightarrow$ are continuous.

Ex. $X \xrightarrow{f} \mathbb{R}^m \xrightarrow{g} \mathbb{R}^p$

If f, g are of class C^k then $g \circ f$ is also of class C^k.

(Induction using chain rule).

Fact: if $f \in C^k(X, \mathbb{R}^m)$, $k \geq 2$ then the partial derivatives of order $\leq k$ are independent of the order of differentiation.

So:

$$\boxed{\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}}$$ for f C^2

$$\left(\longrightarrow \frac{\partial^4 f}{\partial x_1 \partial x_2^3} = \frac{\partial^4 f}{\partial x_2 \partial x_1 \partial x_2^2} = \frac{\partial^4 f}{\partial x_2^3 \partial x_1} \right)$$

f class C^4

$\cdots$

[see script for example 3.55 (2) where the exchange property fails when the second order derivatives are not continuous]

134

<u>Def.</u> $f \in C^2(X, \mathbb{R})$, $x_0 \in X$

$$\underbrace{\mathrm{Hess}_f(x_0) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \right)_{1 \le i,j \le n}}_{\text{symmetric matrix of size } n}$$

"hessian (matrix) of f at x_0"