

Multivariate Calc: gradients, Jacobians, Hessians

Wednesday, September 3, 2025

9:56 AM

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $f(\vec{x}) = \begin{bmatrix} f_1(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix} \in \mathbb{R}^m$

\nwarrow domain \nwarrow co-domain

Def The derivative or Jacobian of f at a point \vec{x} (in the interior of its domain) is the matrix $Df(\vec{x})$ (or sometimes written $J_f(\vec{x})$... or all kinds of variants)

$$Df(\vec{x})_{i,j} = \frac{\partial f_i(\vec{x})}{\partial x_j} \quad \dots \text{if the partial derivatives exist.}$$

$$Df(\vec{x}) \in \mathbb{R}^{m \times n}$$

Special case: $m=1$

Def The gradient of $f: \mathbb{R}^n \rightarrow \mathbb{R}^1$, written $\nabla f(\vec{x}) \in \mathbb{R}^n$, is the transpose of the Jacobian.

$$\text{and } \lim_{\vec{y} \rightarrow \vec{x}} \frac{\| f(\vec{y}) - (f(\vec{x}) + \overbrace{\nabla f(\vec{x})^T \cdot (\vec{y} - \vec{x})}^{\text{1st order Taylor series}}) \|}{\| \vec{y} - \vec{x} \|} = 0$$

Def The Hessian of $f: \mathbb{R}^n \rightarrow \mathbb{R}^1$, written $\nabla^2 f(\vec{x}) \in \mathbb{R}^{n \times n}$, is the Jacobian of the gradient,

$$\nabla^2 f(\vec{x})_{i,j} = \frac{\partial^2 f(\vec{x})}{\partial x_i \partial x_j}$$

Clairaut's Thm says that as long as all these entries are continuous, then $\nabla^2 f(\vec{x})$ is a symmetric matrix, i.e., order of partial derivatives doesn't matter: $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$

Example: $f: \mathbb{R}^2 \rightarrow \mathbb{R}^1$, (in-class exercise)

$$f(x,y) = 4x^3 + 2xy + 13y^2 - 9y^3$$

$$\nabla f(x,y) = \begin{bmatrix} 12x^2 + 2y \\ 2x + 26y - 27y^2 \end{bmatrix},$$

$$\nabla^2 f(x,y) = \begin{bmatrix} 24x & 2 \\ 2 & 26 - 54y \end{bmatrix}$$

\nwarrow i.e. $\frac{\partial^2 f}{\partial x \partial y}$

p. 2: Multivariate Calc

Wednesday, September 3, 2025

10:13 AM

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ i.e. $m=1$ on this page

Directional Derivatives: reducing to 1D case

Def The directional derivative of $f: \mathbb{R}^n \rightarrow \mathbb{R}$ at \vec{x} along \vec{d} is $(\vec{x}, \vec{d} \in \mathbb{R}^n)$

$$f'(\vec{x}; \vec{d}) := \lim_{h \rightarrow 0} \frac{f(\vec{x} + h\vec{d}) - f(\vec{x})}{h} \quad (= \nabla f(\vec{x})^T \cdot \vec{d} = \vec{d}^T \cdot \nabla f(\vec{x}))$$

i.e. the usual 1D derivative of $\varphi(t) = f(\vec{x} + t\vec{d})$

Multivariate Taylor Expansions

$$f(\vec{x}) = \underbrace{f(\vec{x}_0)}_{\text{Scalar}} + \underbrace{\nabla f(\vec{x}_0)^T}_{\text{Vector}} \cdot (\vec{x} - \vec{x}_0) + (\vec{x} - \vec{x}_0)^T \cdot \underbrace{\nabla^2 f(\vec{x}_0)}_{\text{matrix}} (\vec{x} - \vec{x}_0) + \underbrace{O(\|\vec{x} - \vec{x}_0\|^3)}_{\dots \text{ tensors } \dots}$$

We can recover a lot of theorems by reducing to 1D:

Thm Let $\vec{x}, \vec{x}_0 \in \mathbb{R}^n$ and $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be sufficiently smooth, then

$\exists \xi$ on the line segment between \vec{x} and \vec{x}_0 s.t.

$$f(\vec{x}) = f(\vec{x}_0) + \nabla f(\xi)^T \cdot (\vec{x} - \vec{x}_0)$$

Proof

Define $\varphi(t) = f(\vec{x}_0 + t\vec{d})$ where $\vec{d} = \vec{x} - \vec{x}_0$, so $\varphi(0) = f(\vec{x}_0)$ and $\varphi(1) = f(\vec{x})$

Then via Taylor's remainder theorem, $\exists 0 \leq s \leq 1$ s.t.

$$\begin{aligned} f(\vec{x}) &= \varphi(1) = \varphi(0) + \varphi'(s) \cdot (1-0) \\ &= f(\vec{x}_0) + \nabla f(\vec{x}_0 + s\vec{d})^T \cdot \vec{d} \\ &= f(\vec{x}_0) + \nabla f(\xi)^T \cdot (\vec{x} - \vec{x}_0) \quad \square \end{aligned}$$

Chain Rule

Wednesday, September 3, 2025

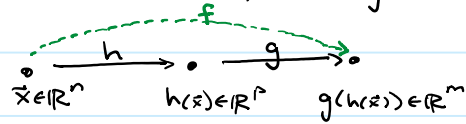
10:14 AM

Scalar Chain Rule $f(x) = g(h(x))$, $f: \mathbb{R}^1 \rightarrow \mathbb{R}^1$
 $f'(x) = g'(f(x)) \cdot h'(x)$

Multivariate only twist: order matters! $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $f(\vec{x}) = g(h(\vec{x}))$

$$J_f(\vec{x}) = \overbrace{J_g(h(\vec{x})) \cdot J_h(\vec{x})}^{m \times n \text{ matrix}}$$

$m \times p \quad p \times n$



$\neq J_h(\vec{x}) \cdot J_g(f(\vec{x})) \leftarrow \text{WRONG!}$

Special case: $m=1$

$$\nabla f(\vec{x})^T = J_f(\vec{x}) = J_g(h(\vec{x})) \cdot J_h(\vec{x}) = \nabla g(h(\vec{x}))^T \cdot J_h(\vec{x})$$

$$\text{ie. } \nabla f(\vec{x}) = J_h(\vec{x})^T \cdot \nabla g(h(\vec{x})) \quad \text{recall } (AB)^T = B^T A^T$$

example from APPM 3310

$$f(\vec{x}) = \frac{1}{2} \|A\vec{x} - \vec{b}\|_2^2 = g(h(\vec{x})) \text{ with } h(\vec{x}) = A\vec{x} - \vec{b}, J_h(\vec{x}) = A$$

$$g(\vec{y}) = \frac{1}{2} \|\vec{y}\|_2^2 = \frac{1}{2} \sum_i y_i^2$$

So via chain rule

$$\nabla f(\vec{x}) = J_h(\vec{x})^T \cdot \nabla g(h(\vec{x}))$$

$$\nabla g(\vec{y}) = \vec{y}$$

$$= A^T \cdot (A\vec{x} - \vec{b})$$

thus setting $\nabla f(\vec{x}) = 0$ gives rise to $A^T(A\vec{x} - \vec{b}) = 0$,
the normal equations!

Mathematical Formalities (optional, not covered in class)

Wednesday, September 3, 2025 10:13 AM

$$f: \mathbb{R}^n \rightarrow \mathbb{R} \quad \text{i.e. } m=1 \text{ on this page}$$

Differentiability in \mathbb{R}^n $n \geq 1$

There are different notions of differentiability. For \mathbb{R}^1 , these all coincide luckily

1) (weakest) Partial derivatives exist, i.e., directional derivatives along coordinate axes

$$\text{i.e., } \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \text{ all exist}$$

$$\text{ex, } \mathbb{R}^2, f(x,y) = (xy)^{1/3}, \quad \frac{\partial f}{\partial x} = \frac{1}{3} x^{-2/3} y^{1/3} \text{ and } \frac{\partial f}{\partial y} \text{ also exists}$$

but along line $y=x$, let $g(x) = f(x,x) = x^{2/3}$, not differentiable at 0 since $g'(x) = \frac{2}{3} x^{-1/3}$

2) (next weakest) Gâteaux differentiable, i.e., directional derivatives exist for all directions
version 1

$$\text{i.e., } \forall \text{ directions } d \in \mathbb{R}^n, \quad f'(x; d) := \lim_{h \rightarrow 0} \frac{f(x+hd) - f(x)}{h} \text{ exists.}$$

2') (next weakest) Gâteaux diff, version 2 (authors don't agree) i.e. Gradient Exists
same as 2) but also require $d \mapsto f'(x; d)$ is a bounded linear function "∇f(x)"

Saying it's linear means, in a Hilbert space (i.e., using Riesz \uparrow in \mathbb{R}^n , this comes for free)

$$\text{we can write } f'(x; d) = \boxed{\langle \nabla f(x), d \rangle} \quad \text{COMMON NOTATION}$$

3) (strictest) Fréchet differentiable

means $d \mapsto f'(x; d)$ is a linear function (like 2')

and there's a uniform rate of convergence (in "h") independent of the direction,

$$\text{i.e., } \lim_{\|d\| \rightarrow 0} \frac{\| (f(x) + \langle \nabla f(x), d \rangle) - f(x+d) \|}{\|d\|} = 0$$

in case $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m > 1$

4) (even stricter than strict)

$$f \in C^1 \quad \text{i.e., } \nabla f(x) \text{ exists } \forall x \text{ and it's continuous}$$

This implies Fréchet (hence Gâteaux) diff. *

[So for simplicity, we usually assume $f \in C^1$
and don't worry about the details]

* I'm pretty sure but not 100%... it's not obvious.

(In particular, we often assume ∇f is Lipschitz continuous,
even stronger assumption than $f \in C^1$!)