E600 Mathematics

Chapter 3: Multivariate Calculus

Martin Reinhard

August 27, 2021

This chapter discusses

- A formal introduction to multi-dimensional functions
- Key function properties: invertability, convexity (and concavity)
- Multivariate differentiation (main focus)
  - Formal definition and derivation
  - Application
- Multivariate integration: concept and key theorems

- Thus far: Linear Algebra (linear operations, equation systems)
- Now: analysis of functions, study of (small) variations
- Here: generalizing the derivative to functions $f : \mathbb{R}^n \mapsto \mathbb{R}^m$
- Why?: Optimization problems with many variables (goods, production inputs, statistical parameters)
- Many struggles in the 1st PhD semester were encountered because of issues with understanding derivatives. . .

- Function $f : X \mapsto Y$ with domain X, codomain Y and image $\text{im}(f) = f[X]$
  - $X \subseteq \mathbb{R}$: univariate function
  - $X \subseteq \mathbb{R}^n$: multivariate function
  - $Y \subseteq \mathbb{R}$: real-valued function
  - $Y \subseteq \mathbb{R}^m$: vector-valued function
  - How to call $f : \mathbb{R}^3 \mapsto \mathbb{R}^2$?
- Examples:
  - Multivariate, real-valued function: $x \mapsto \|x\|$, $x \mapsto x'Ax$, $(x, y) \mapsto x \cdot y$
  - Multivariate, vector-valued function: $x \mapsto Ax$
- Graph:

$$G(f) = \{(x, y) \in X \times Y : y = f(x)\} = \{(x, f(x)) : x \in X\}$$

- Inverse function f^{-1} of f: $f(f^{-1}(y)) = y$ and $f^{-1}(f(x)) = x$
- Ch. 0: For $X, Y \subseteq \mathbb{R}$, we can *invert* $f : X \mapsto Y$ if, and only if, for every $y \in Y$ we have exactly one $x(y) \in X$ so that $f(x(y)) = y$
- The two conditions transfer to arbitrary X, Y: for every $y \in Y$, ...
    - at least one x is mapped onto y: $\exists x \in X : f(x) = y$ ("f is surjective/onto")
    - no more than one x is mapped onto y: $(x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2))$ ("f is injective/one-to-one")
- If f is both injective and surjective, we call f *bijective*
- f invertible $\Leftrightarrow$ f bijective ((simple) proof in script)

- Note that $f^{-1} \circ f : X \mapsto X, x \mapsto x$ and $f \circ f^{-1} : Y \mapsto Y, y \mapsto y$
  - $f^{-1} \circ f$ and $f \circ f^{-1}$ are equal to *identity functions*, the *identity object* in function spaces
  - Function inversion follows our standard inversion logic "$x^{-1}x = 1$"!
- What's the difference between f^{-1}, $f^{-1}(y)$, $f^{-1}[y]$ and $f^{-1}[\{y\}]$?

- Note that $f^{-1} \circ f : X \mapsto X, x \mapsto x$ and $f \circ f^{-1} : Y \mapsto Y, y \mapsto y$
  - $f^{-1} \circ f$ and $f \circ f^{-1}$ are equal to *identity functions*, the *identity object* in function spaces
  - Function inversion follows our standard inversion logic "$x^{-1}x = 1$"!
- What's the difference between f^{-1}, $f^{-1}(y)$, $f^{-1}[y]$ and $f^{-1}[\{y\}]$?

. . . Pre-images and inverse functions are fundamentally different objects!!!

Definition (Convex and Concave Real Valued Function)

Let $X \subseteq \mathbb{R}^n$ be a convex set. A function $f : X \to \mathbb{R}$ is convex if for any $x, y \in X$ and $\lambda \in [0,1]$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

Moreover, if for any $x, y \in X$ such that $y \neq x$ and $\lambda \in (0,1)$,

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

we say that f is strictly convex. Moreover, we say that f is (strictly) concave if $-f$ is (strictly) convex.

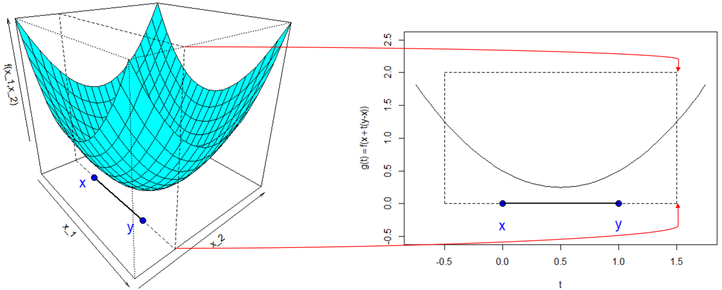
Alternative characterization of concavity (line 1) and strict concavity (line 2)

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y) \qquad \forall x, y \in X \forall \lambda \in [0,1],$$

$$f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda) f(y) \qquad \forall x, y \in X \text{ so that } x \neq y \text{ and } \forall \lambda \in (0,1).$$

- In what follows: focus on convexity
- Recall: $\lambda x + (1-\lambda)y$ ($\lambda \in [0,1]$) is a convex combination of x and y
    - Convexity of functions = statement about convex combinations across domain and codomain of the function!
    - $f(\lambda x + (1-\lambda)y)$ must always be well-defined $\rightarrow$ convex domain
- $G(f) \subseteq \mathbb{R}^2$ (i.e. f univariate, real-valued function):
    - $(1-\lambda)x + \lambda y$, $\lambda \in [0,1]$ defines an interval between x and y
    - $(1-\lambda)f(x) + \lambda f(y)$ is the line piece connecting $f(x)$ and $f(y)$

      (you'll shortly see why we use $1-\lambda$ as the coefficient of $x\ldots$)
    - Let's draw a convex and a concave function



- $f : X \mapsto \mathbb{R}$, $X \subseteq \mathbb{R}^2$
- For any fixed $x, y \in X$, $\lambda y + (1-\lambda)x = x + \lambda(y-x)$ expands in a single direction

- Gives univariate function $t \mapsto f(x + t(y-x))$

$\Rightarrow$ convex?

- Also for $X \subseteq \mathbb{R}^n$: fixing $x, y \in X$ reduces convexity to one dimension

$\rightarrow$ f is convex if and only if any univariate reduction is convex

- After picking $x \in X$, choosing $y \in X$ arbitrarily is equivalent to choosing $z \in \mathbb{R}^n$ with $x + z \in X$ arbitrarily ($z = y - x$). This gives:

Theorem (Graphical Characterization of Convexity)

Let $X \subseteq \mathbb{R}^n$ be a **convex set** *and $f : X \mapsto \mathbb{R}$. Then, f is (strictly) convex if and only if $\forall x \in X$ and $\forall z \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ with $x + z \in X$, the function $g : \mathbb{R} \mapsto \mathbb{R}, t \mapsto f(x + tz)$ is (strictly) convex.*

Let's use this theorem ("strictly" variant) to give an example of a convexity proof.

Corollary (Disproving Convexity)

Let $X \subseteq \mathbb{R}^n$ be a **convex set** *and $f : X \mapsto \mathbb{R}$. Then, if there exist $x_0 \in X$ and $i \in \{1, \ldots, n\}$ such that $g : \mathbb{R} \mapsto \mathbb{R}, t \mapsto f(t \mapsto x + t \cdot e_i)$ is not (strictly) convex, then f is not (strictly) convex.*

- Necessary condition of convexity: convex in every *fundamental direction* of $\mathbb{R}^n$
- Consider

$$f(x) = h(x_1, \ldots, x_{n-1}) \cdot \sqrt{x_n}$$

where h is an arbitrarily complex, unspecified function. Is f convex?

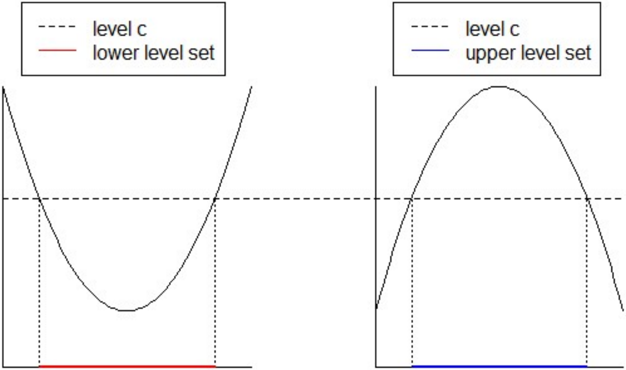
- Optimization: convexity immensely helpful, but restrictive concept
- Can we weaken the concept and preserve (most of!) the desirable properties? Yes!
- Level sets in the domain of f:

Definition (Lower and Upper Level Set of a Function)

Let $X \subseteq \mathbb{R}^n$ be a convex set and $f : X \to \mathbb{R}$ be a real-valued function. Then, for $c \in \mathbb{R}$, the sets

$$L_c^- := \{x \in X : f(x) \leq c\} \quad \text{and} \quad L_c^+ := \{x \in X : f(x) \geq c\}$$

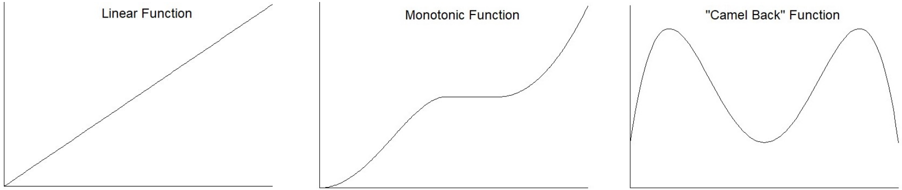
are called the lower-level and upper level set of f at c, respectively.



Quasi-convex (-concave) function: *any* lower (upper) level set convex

2. Basics: Invertability and Convexity

Level Sets of Quasi-Convex and -Concave Functions



Which functions are quasi-convex/quasi-concave? Which are *quasi-linear*?

- Quasi-linear: both quasi-convex and quasi-concave
- Motivation: only linear functions are both convex and concave

Theorem (Quasiconvexity, Quasiconcavity)

Let $X \subseteq \mathbb{R}^n$ be a convex set. A real-valued function $f : X \to \mathbb{R}$ is quasiconvex if and only if

$$\forall x, y \in X \forall \lambda \in [0,1] : f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$$

Conversely, f is quasiconcave if and only if

$$\forall x, y \in X \forall \lambda \in [0,1] : (\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}$$

Analogous *definition*: Strict Quasiconvexity (line 1), Strict Quasiconcavity (line 2)

$$\forall x, y \in X \text{ such that } x \neq y \text{ and } \forall \lambda \in (0,1) : f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$$

$$\forall x, y \in X \text{ such that } x \neq y \text{ and } \forall \lambda \in (0,1) : f(\lambda x + (1-\lambda)y) > \min\{f(x), f(y)\}$$

- What is (multivariate) Calculus? Defininition Wikipedia (summarized)
    - “Mathematical study of continuous change”
    - Differential calculus: instantaneous/marginal rates of change and slopes of curves
    - Integral calculus: accumulation of quantities, areas under and between curves
    - Fundamental theorem of calculus: integration and differentiation are *inverse operations* (intuition?)
- Why care?
    - Cannot optimize without derivatives
    - Economics: marginal utility, accumulations across households

- As before: start from what we know and generalize
- If $X \subseteq \mathbb{R}$, what is "the slope" of $f : X \mapsto \mathbb{R}$?
- Relative change of $f(x)$ given variation in x at $x_0 \in X$:

$$\frac{\Delta f(x)}{\Delta x} := \frac{f(x) - f(x_0)}{x - x_0} = \frac{f(x_0 + h) - f(x_0)}{h}, \qquad h := x - x_0 \in \mathbb{R}$$

- Marginal/instantaneous rate of change: limit $h \to 0$ (existence?)

Definition (Univariate Real-Valued Function: Differentiability and Derivative)

Let $X \subseteq \mathbb{R}$ and consider the function $f : X \mapsto \mathbb{R}$. Let $x_0 \in X$. If

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists, f is said to be differentiable at x_0, and we call this limit the **derivative of f at** x_0, *denoted by $f'(x_0)$. If for all $x_0 \in X$, f is differentiable at x_0, f is said to be differentiable over X or differentiable. Then, the function $f' : X \mapsto \mathbb{R}, x \mapsto f'(x)$ is called the* **derivative** *of f.*

- Differentiability: point-specific vs. domain (e.g. $|\cdot|$)
- Derivative f' = function, derivative at x_0, $f'(x_0)$ = real number!

Definition (Univariate Real-Valued Functions: Differential Operator)

Let $X \subseteq \mathbb{R}$*, define* $D^1(X,\mathbb{R}) = \{f : X \mapsto \mathbb{R} : f \text{ is differentiable over } X\}$*, and let* $F_X := \{f : X \mapsto \mathbb{R}\}$*. Then, the differential operator is defined as the function*

$$\frac{d}{dx} : D^1(X,\mathbb{R}) \mapsto F_X, f \mapsto f'$$

where f' *denotes the derivative of* $f \in D^1(X,\mathbb{R})$*.*

- (Differential) Operator: function between function spaces
- $f' = \frac{d}{dx}(f)$ is a specific value in the codomain of $\frac{d}{dx}$ (just like $f'(x)$)
- Formally precise $f'(x) = \left[\frac{d}{dx}(f)\right](x)$ vs. convention: $f'(x) = \frac{df}{dx}(x)$
- Please don't write $\frac{df(x)}{dx}$

- Levels of objects in differentiation: operator, function, value
- Let's practice this distinction with some common rules
- Recall: basis operations in function spaces
    - "+": $(f+g)(x) = f(x) + g(x)$
    - "·": $(\lambda f)(x) = \lambda \cdot f(x)$
- Function product $(fg)(x) = f(x) \cdot g(x)$
- quotient in analogy if $\forall x \in X : g(x) \neq 0$

Theorem (Rules for Univariate Derivatives)

Let $X \subseteq \mathbb{R}$, $f, g \in D^1(X, \mathbb{R})$ and $\lambda, \mu \in \mathbb{R}$. Then,

(i) *(Linearity) $\lambda f + \mu g$ is differentiable and $\frac{d}{dx}(\lambda f + \mu g) = \lambda \frac{df}{dx} + \mu \frac{dg}{dx}$,*

(ii) *(Product Rule) The product fg is differentiable and $\frac{d}{dx}(fg) = \frac{df}{dx} \cdot g + f \cdot \frac{dg}{dx}$*

(iii) *(Quotient Rule) If $\forall x \in X$, $g(x) \neq 0$, the quotient f/g is differentiable and $\frac{d}{dx}(f/g) = \frac{\frac{df}{dx} \cdot g - f \cdot \frac{dg}{dx}}{g \cdot g}$*

(iv) *(Chain Rule) if $g \circ f$ exists, the function is differentiable and $\frac{d}{dx}(g \circ f) = \left(\frac{dg}{dx} \circ f\right) \cdot \frac{df}{dx}$.*

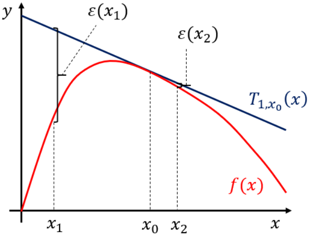
Script: rules for specific values and differentiability at $x_0 \in X$

- Now: local behavior of multivariate functions we cannot sketch?
- For univariate, real-valued functions, differentiability of f at $x_0 \in X$ implies...

1. **continuity** at x_0
2. **Existence of a good linear approximation** to f around x_0

and differentiability of f on $(a, b) \subseteq \mathbb{R}$ implies that

3. the **sign** of f' is determines if the function is increasing, decreasing, or constant



- Taylor = key take-away from this class!
- First order Taylor approximation to f at x_0:

$$T_{1,x_0}(x) = f(x_0) + f'(x_0)(x - x_0)$$

- Error: $\varepsilon_{1,x_0}(x) := f(x) - T_{1,x_0}(x)$ (formula: next slide)

- "Good" approximation: $\lim_{x \to x_0} \frac{\varepsilon_1(x)}{x - x_0} = 0$ (intuition?; caution?)
- Taylor *expansion* of first order: decomposition of f into linear and non-linear term, i.e.

$$f(x) = T_{1,x_0}(x) + \varepsilon_{1,x_0}(x)$$

- *Expansion* includes the error, *approximation* does not

Theorem (Taylor Expansion for Univariate Functions)

Let $X \subseteq \mathbb{R}$ *and* $f \in D^d(X, \mathbb{R})$ *where* $d \in \mathbb{N} \cup \{\infty\}$. *For* $N \in \mathbb{N} \cup \{\infty\}$, $N \leq k$, *the Taylor expansion of order* N *for* f *at* $x_0 \in X$ *is*

$$f(x) = T_{N,x_0}(x) + \varepsilon_{N,x_0}(x) = f(x_0) + \sum_{n=1}^{N} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \varepsilon_{N,x_0}(x),$$

where $\varepsilon_{N,x_0}(x)$ *is the approximation error of* T_{N,x_0} *for* f *at* $x \in X$. *Then, the approximation quality satisfies* $\lim_{h \to 0} \varepsilon_{N,x_0}(x_0 + h)/h^N = 0$. *Further, if* f *is* $N + 1$ *times differentiable, there exists a* $\lambda \in (0, 1)$ *such that*

$$\varepsilon_{N,x_0}(x_0 + h) = \frac{f^{(N+1)}(x_0 + \lambda h)}{(N + 1)!} h^{N+1}.$$

- Faculty of $n \in \mathbb{N}$: $n! = 1 \cdot 2 \cdot \ldots \cdot (n - 1) \cdot n$

- Approximation quality: $\lim_{h \to 0} \varepsilon_N(x_0 + h)/h^N = 0$
  - The larger N, the "faster" $h^N \to 0$ (think 0.1^n for increasing n)
  - Larger N increase order of approximation quality
  - Script gives proof for $N = 1, 2$, general intuition is similar
- Mean Value Theorem (corollary of Taylor's theorem): for any differentiable $f : X \mapsto \mathbb{R}$ ($X \subseteq \mathbb{R}$), for any $x_1, x_2 \in X$ such that $x_2 > x_1$, there exists $x^* \in (x_1, x_2)$ such that

$$f'(x^*) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

  - Useful to check if *critical values* ($f'(x) = 0$) exist
  - Proof: In-class exercises

- Roadmap for multivariate derivatives ($f : X \mapsto Y$, esp. $X \subseteq \mathbb{R}^n$)
  1. How to formally think about a multivariate derivative?
     - derivative should describe expansion in *any possible* direction
     - $X \subseteq \mathbb{R}$: variation on an infinitely small intervall/ball around x_0
  2. Does an intuitively plausible candidate meet the formal definition?
- Recall: convergence
  - univariate: $\lim_{x \to 0} f(x) = c$: $|f(x) - c| < \varepsilon$ for $|x - 0| = |x| < \delta$
  - multivariate: $\lim_{x \to 0} f(x) = c$: $|f(x) - c| < \varepsilon$ for $\|x\| < \delta$

$\rightarrow$ tells us how to think about "$\lim_{h \to 0}$" more generally!

- when $n = 1$, d^* is the derivative of f at x_0 if

$$d^* = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

- Problem: if $n > 1$, the denominator has a *vector*; not defined
- But: expression is *equivalent* to (let's see why)

$$\lim_{h \to 0} \frac{|f(x_0 + h) - f(x_0) - d^* \cdot h|}{\|h\|} = 0$$

where $\| \cdot \|$ is a norm on $\mathbb{R}$; and norms generalize to $\mathbb{R}^n$!

4. Multivariate Calculus: Differentiation

Definition (Multivariate Derivative of Real-valued Functions)

Let $X \subseteq \mathbb{R}^n$ and $f : X \mapsto \mathbb{R}$. Further, let $x_0 \in int(X)$ (interior point). Then, f is differentiable at x_0 if there exists $d^* \in \mathbb{R}^{1\times n}$ such that

$$\lim_{\|h\|\to 0} \frac{|f(x_0+h) - f(x_0) - d^*h|}{\|h\|} = 0.$$

Then, we call d^* the derivative of f at x_0, denoted $\frac{df}{dx}(x_0)$ or $D_f(x_0)$. If f is differentiable at any $x_0 \in X$, we say that f is differentiable, and we call $\frac{df}{dx} : X \mapsto \mathbb{R}, x \mapsto \frac{df}{dx}(x)$ the derivative of f.

- Interior point: able to consider balls around x_0 on which f is defined
- Most textbooks use $D_f(x_0)$ rather than $\frac{df}{dx}(x_0)$, is the same thing!
- Derivative operator as before: mapping between function spaces

4. Multivariate Calculus: Differentiation

Definition (Multivariate Derivative of Vector-valued Functions)

Let $X \subseteq \mathbb{R}^n$ *and* $f : X \mapsto \mathbb{R}^m$*. Further, let* $x_0 \in int(X)$ *(interior point). Denote* ${}_k\|\cdot\|$ *as a norm of* $\mathbb{R}^k$*,* $k \in \{n, m\}$*. Then,* f *is differentiable at* x_0 *if there exists a matrix* $D^* \in \mathbb{R}^{m\times n}$ *such that*

$$\lim_{{}_n\|h\|\to 0} \frac{{}_m\|f(x_0+h) - f(x_0) - D^*h\|}{{}_n\|h\|} = 0,$$

Then, we call D^* *the derivative of* f *at* x_0*, denoted* $\frac{df}{dx}(x_0)$ *or* $D_f(x_0)$*. If* f *is differentiable at any* $x_0 \in X$*, we say that* f *is differentiable, and we call* $\frac{df}{dx} : X \mapsto \mathbb{R}^{m\times n}, x \mapsto \frac{df}{dx}(x)$ *the derivative of* f*.*

- Numerator norm: codomain, denominator norm: domain
- Derivative as matrix: D^*h must be vector of same length as $f(x)$
- Actually: encompasses the previous definition $(m = 1)$

- Roadmap for multivariate derivatives
  - ✓ How to formally think about a multivariate derivative?
  - 2. Does an intuitively plausible candidate meet the formal definition?
- Idea:
  - For $n = 1$, $\frac{df}{dx}(x_0)$ is scalar and characterizes the instantaneous change along the one axis (i.e., fundamental direction) of $\mathbb{R}$
  - For $n > 1$, $\frac{df}{dx}(x_0)$ is a vector of length n $\rightarrow$ collection of instantaneous changes along all n individual axes of $\mathbb{R}^n$?

$\Rightarrow$ Directional derivative: expansion of f from x_0 in a direction $z \neq \mathbf{0}$

- Directional derivative: let $f_{z,x_0} : \mathbb{R} \mapsto \mathbb{R}, t \mapsto f(x_0 + tz)$ for $z \neq \mathbf{0}$
  - Univariate directional derivative of f in direction z at x_0: $\frac{df_{z,x_0}}{dt}(0)$
- Partial derivative of f at x_0 with respect to x_j (∂ vs. d!):

$$\frac{\partial f}{\partial x_j}(x_0) = \frac{df_{e_j,x_0}}{dt}(0) = \frac{d}{dt} f(x_0 + te_j)|_{t=0}$$
$$= \frac{d}{dt}[f(x_{0,1}, \dots, x_{0,j-1}, \mathbf{x_{0,j}} + \mathbf{t}, x_{0,j+1}, \dots x_{0,n})]\Big|_{t=0}$$

  - Variation in direction j around x_0 ("holding x_l, $l \neq j$ constant")
  - Also: j-th partial derivative (of f at x_0); sometimes denoted $f_j(x_0)$
- Gradient: *ordered* collection of partial derivatives (**row** vector!)

$$\nabla f(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \frac{\partial f}{\partial x_2}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0) \right)$$

- Partial differentiability
  - $f : X \mapsto \mathbb{R}$ partially differentiable (p.d.) at x_0: all partial derivatives $\frac{\partial f}{\partial x_j}(x_0)$ and therefore the gradient at $x_0 \in X$, $\nabla f(x_0)$, exists
  - "point-specific to general": $f : X \mapsto \mathbb{R}$ p.d.: f p.d. at any $x_0 \in X$
  - Set of p.d. functions from X to $\mathbb{R}$: $D_p^1(X, \mathbb{R}) = \{f : X \mapsto \mathbb{R} : f \text{ is p.d.}\}$
- Recall: univariate derivative is a real-valued function
  - $\frac{\partial f}{\partial x_j} : X \mapsto \mathbb{R}$, $x_0 \mapsto \frac{\partial f}{\partial x_j}(x_0)$ is a real-valued function
  - $\nabla f : X \mapsto \mathbb{R}^{1\times n}$, $x_0 \mapsto \nabla f(x_0)$ is a (real row-)vector-valued function
- associated operators: mappings between function spaces
  - $\frac{\partial}{\partial x_j} : D_p^1(X, \mathbb{R}) \mapsto F_X, f \mapsto f_j = \frac{\partial f}{\partial x_j}$
  - $\nabla : D_p^1(X, \mathbb{R}) \mapsto F_X^{1\times n}, f \mapsto \nabla f$

Consider the following functions $\mathbb{R}^2 \mapsto \mathbb{R}$:

- $f^1(x_1, x_2) = x_1 + x_2$
- $f^2(x_1, x_2) = x_1 x_2$
- $f^3(x_1, x_2) = x_1 x_2^2 + \cos(x_1)$

Consider an arbitrary point $x_0 = x \in \mathbb{R}$. Compute the gradients of f^1, f^2 and f^3 at x_0!

How do the partial derivatives depend on the location x?

Now for the actual derivative: can we use the gradient?

Theorem (The Gradient and the Derivative)

Let $X \subseteq \mathbb{R}^n$ and $f : X \mapsto \mathbb{R}$ such that f is differentiable at $x_0 \in int(X)$. Then, all partial derivatives of f at x_0 exist, and $\frac{df}{dx}(x_0) = \nabla f(x_0)$.

- Verbally: "derivative exists $\Rightarrow$ derivative $=$ gradient"; what about $\Leftarrow$?

Theorem (Partial Differentiablility and Differentiability)

Let $X \subseteq \mathbb{R}^n$, $f : X \mapsto \mathbb{R}$ and $x_0 \in int(X)$. If all the partial derivatives of f at x_0 exist and are continuous, then f is differentiable.

- Set of continuously differentiable functions:

$$C^1(X, \mathbb{R}) := \left\{ f : X \mapsto \mathbb{R} : \left(\forall j \in \{1, \ldots, n\} : \frac{\partial f}{\partial x_j} \text{ is continuous} \right) \right\}$$

- $f \in C^1(X, \mathbb{R}) \Rightarrow f$ is differentiable

- Partial differentiability and differentiability
    - Generally, if f is differentiable, the derivative is equal to the gradient
    - $\Rightarrow$ If the gradient does not exist, f is not differentiable
    - Theoretically: may encounter weird D^1 but not C^1 functions; issue not too relevant in (economic) practice
- In applications: taking the derivative of $f : X \mapsto \mathbb{R}$, $X \subseteq \mathbb{R}^n$
    1. Compute the gradient ∇f (if it exists)
    2. Are all partial derivatives continuous? If so: ∇f is the derivative!
- What about $f : X \mapsto \mathbb{R}^m$?

- Consider $X \subseteq \mathbb{R}^n$, $f : X \mapsto \mathbb{R}^m$
- f is ordered collection of real-valued functions which we already know how to handle:

$$f = \begin{pmatrix} f^1 \\ f^2 \\ \vdots \\ f^m \end{pmatrix} \text{ so that } \forall x \in X : f(x) = \begin{pmatrix} f^1(x) \\ f^2(x) \\ \vdots \\ f^m(x) \end{pmatrix}$$

where for any $i \in \{1, \ldots, m\}$, $f^i : X \mapsto \mathbb{R}$ (example?)

- Idea: ordered collection of derivatives, i.e.

$$\frac{df}{dx} = \begin{pmatrix} \nabla f^1 \\ \nabla f^2 \\ \vdots \\ \nabla f^m \end{pmatrix}$$

4. Multivariate Calculus: Differentiation

Vector-valued Functions 2/3

Definition (Jacobian)

Let $n, m \in \mathbb{R}^n$, $X \subseteq \mathbb{R}^n$ and $f : X \mapsto \mathbb{R}^m$ and for $i \in \{1, \ldots, m\}$, let $f^i : \mathbb{R}^n \mapsto \mathbb{R}$ such that $f = (f^1, \ldots, f^m)'$. Let $x_0 \in X$. Then, if at x_0, $\forall i \in \{1, \ldots, m\}$, f^i is partially differentiable with respect to any x_j, $j \in \{1, \ldots, n\}$, we call

$$J_f(x_0) = \begin{pmatrix} \nabla f^1(x_0) \\ \nabla f^2(x_0) \\ \vdots \\ \nabla f^m(x_0) \end{pmatrix} = \begin{pmatrix} f^1_1(x_0) & f^1_2(x_0) & \ldots & f^1_n(x_0) \\ f^2_1(x_0) & f^2_2(x_0) & \ldots & f^2_n(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ f^m_1(x_0) & f^m_2(x_0) & \ldots & f^m_n(x_0) \end{pmatrix}$$

the Jacobian of f at x_0. If the above holds at any $x_0 \in X$, we call the mapping $J_f : \mathbb{R}^n \mapsto \mathbb{R}^{n \times m}, x_0 \mapsto J_f(x_0)$ the Jacobian of f.

- All partial derivative of any f^i must exist (we write $f \in D^1_p(X, \mathbb{R}^n)$)

- Jacobian collects expansion in all fundamental directions of all sub-functions f^i, $i \in \{1, \ldots, m\}$. $\rightarrow$ Jacobian = derivative?

Theorem (The Jacobian and the Derivative)

Let $X \subseteq \mathbb{R}^n$, $f : X \mapsto \mathbb{R}^m$ and $f^1, \ldots, f^m : X \mapsto \mathbb{R}$ such that $f = (f^1, \ldots, f^m)'$. Further, let $x_0 \in int(X)$ (interior point), and suppose that f is differentiable at x_0. Then, for any f^i, $i \in \{1, \ldots, m\}$, all partial derivatives of f^i at x_0 exist, and $\frac{df}{dx}(x_0) = J_f(x_0)$.

- As before: derivative exists if all partial deriv's are continuous

- Why did our intuitive conjecture correspond to the derivative?
- Recall lecture 1...
  - Vector spaces: generalize key intuitions of lower-dimensional spaces
  - Minimal structure (addition and multiplication by a constant)...
  - ...and an *axiomatic* way of thinking about distances
  - ...was **all** we needed to generalize a complex and important concept such as function differentiation

Theorem (Rules for Multivariate Derivatives)

Let $X \subseteq \mathbb{R}^n$, $f, g : X \mapsto \mathbb{R}^m$ and $h : \mathbb{R}^m \mapsto \mathbb{R}^k$. Suppose that f, g and h are differentiable functions. Then,

(i) *(Linearity) For all $\lambda, \mu \in \mathbb{R}$, $\lambda f + \mu g$ is differentiable and* $\frac{d(\lambda f + \mu g)}{dx} = \lambda \frac{df}{dx} + \mu \frac{dg}{dx}$.

(ii) *(Product Rule) $f' \cdot g$ is differentiable and* $\frac{d(f' g)}{dx} = f' \cdot \frac{dg}{dx} + g' \cdot \frac{df}{dx}$.

(iii) *(Chain Rule) $h \circ f$ is differentiable and* $\frac{d(h \circ f)}{dx} = (\frac{dh}{dx} \circ f) \cdot \frac{df}{dx}$.

- Product rule: f', g' = *transpose*, not derivative; Quotient rule?
- Careful about order (matrix products are not commutative)!
- CR variant: for $f(g(x)) = f(y(x), x)$ (L: precise; R: convention):

$$\frac{df \circ g}{dx} = \frac{\partial f \circ g}{\partial y}\frac{dy}{dx} + \frac{\partial f \circ g}{\partial x} \quad \text{vs.} \quad \frac{df}{dx} = \frac{\partial f}{\partial y}\frac{dy}{dx} + \frac{\partial f}{\partial x}$$

- Thus far: first derivative operator $(\cdot)'$ generalized to ∇/J
- In univariate, real-valued case: $f'' = (f')'$, we can generalize this logic
- Recall: derivative increases order in codomain
  - Derivative of $f : \mathbb{R}^n \mapsto \mathbb{R}$ is vector-valued: $\nabla f : \mathbb{R}^n \mapsto \mathbb{R}^{1\times n}$
  - Derivative of $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ is matrix-valued: $J_f : \mathbb{R}^n \mapsto \mathbb{R}^{m\times n}$
  - Derivative of Jacobian?

  ... Let's focus on real-valued functions to avoid the third dimension
- Expectation: first derivative is vector $\rightarrow$ second is matrix
  - First derivative = gradient: $\nabla f : \mathbb{R}^n \mapsto \mathbb{R}^{1\times n}$
  - Second derivative = derivative of transposed gradient: $\frac{d}{dx}(\nabla f)'$

- If $\frac{\partial f}{\partial x_i}$ is differentiable at x_0, the (i, j)-second order partial derivative at x_0 is

$$f_{i,j}(x_0) = \frac{\partial f_i}{\partial x_j}(x_0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)$$

Definition (Hessian or Hessian Matrix)

Let $X \subseteq \mathbb{R}^n$ be an open set and $f : X \mapsto \mathbb{R}$. Further, let $x_0 \in X$, and suppose that f is differentiable at x_0 and that all second order partial derivatives of f at x_0 exist. Then, the Hessian of f at x_0 is the matrix

$$H_f(x_0) = \begin{pmatrix} \nabla f_1(x_0) \\ \nabla f_2(x_0) \\ \vdots \\ \nabla f_n(x_0) \end{pmatrix} = \begin{pmatrix} f_{1,1}(x_0) & f_{1,2}(x_0) & \cdots & f_{1,n}(x_0) \\ f_{2,1}(x_0) & f_{2,2}(x_0) & \cdots & f_{2,n}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ f_{n,1}(x_0) & f_{n,2}(x_0) & \cdots & f_{n,n}(x_0) \end{pmatrix}$$

- If $(\nabla f)'$ is differentiable, we already know that $\frac{d}{dx}(\nabla f)' = H_f$!

- Let $C^k(X) = C^k(X, \mathbb{R})$ (codomain $\mathbb{R}$ as implicit second argument):

$$C^k(X) = \{f : X \mapsto \mathbb{R} : \text{All } k\text{-th order part. deriv's are continuous}\}$$

Theorem (Schwarz's Theorem/Young's Theorem)

Let $X \subseteq \mathbb{R}^n$ be an open set and $f : \mathbb{R}^n \mapsto \mathbb{R}$. If $f \in C^k(X)$, then the order in which derivatives up to order k are taken can be permuted.

- If $f \in C^2(X)$, then
    - $\nabla f \in C^1(X) \Rightarrow$ differentiable, and
    - derivative = Hessian is symmetric!

Corollary (Hessian and Gradient)

Let $X \subseteq \mathbb{R}^n$ and $f \in C^2(X)$. Then, the Hessian is symmetric and corresponds to the Jacobian of the transposed gradient: $H_f = J_{(\nabla f)'}$.

Let $f(x_1, x_2, x_3) = x_1 x_2^2 + \cos(x_1)e^{x_3}$. Is f twice differentiable? If so, compute the second derivative!

Theorem (Second Order Multivariate Taylor Approximation)

Let $X \subseteq \mathbb{R}^n$ be an open set and consider $f \in C^2(X)$. Let $x_0 \in X$. Then, the second order Taylor approximation to f at $x_0 \in X$ is

$$T_{2,x_0}(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + \frac{1}{2}(x - x_0)' \cdot H_f(x_0) \cdot (x - x_0).$$

The error $\varepsilon_{2,x_0}(x) = f(x) - T_{2,x_0}(x)$ approaches 0 at a faster rate than $\|x - x_0\|^2$, i.e. $\lim_{\|h\|\to 0} \frac{\varepsilon_{2,x_0}(x+h)}{\|h\|^2} = 0$.

- Zero and first order approximation in analogy
- Error formula for first order: there exists $\lambda \in (0, 1)$ so that

$$\varepsilon_{1,x_0}(x_0 + h) = \frac{1}{2}h' \cdot H_f(x_0 + \lambda h) \cdot h$$

- Taylor expansion like before

- Directional derivative of f at x_0 in direction $z \neq \mathbf{0}$ (Chain Rule):

$$\frac{d}{dt}f(x_0 + tz)\Big|_{t=0} = \nabla f(x_0) \cdot z = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0) \cdot z_i$$

- Notation: $z = (dx_1, \ldots, dx_n)$ as vector of *relative variation* in the arguments; df as resulting *relative* induced marginal change

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i; \qquad df(x_0) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0) dx_i$$

- In economics:
  - variation in fixed ratios/specific directions $\rightarrow$ comparative statics
  - Consideration is *relative*: fix one reference variable j with $dx_j = 1$
  - Concerns *marginal* variation; do not consider fixed, non-zero changes!

- For $X \subseteq \mathbb{R}$, $f \in C^2(X)$ (proof in script):
  1. f is convex if and only if $\forall x \in X : f''(x) \geq 0$ (equivalent condition)
  2. If $\forall x \in X : f''(x) > 0$, then f is strictly convex (sufficient condition)
- Recall: we can study $g : \mathbb{R} \mapsto \mathbb{R}, t \mapsto f(x + tz)$ for $x, z \in \mathbb{R}^n$, $z \neq \mathbf{0}$
  - If $f \in C^2(X)$ then especially $g \in C^2(\mathbb{R})$ for fixed x, z
  - Second derivative (chain rule; cf. directional derivative):

$$g''(t) = z' H_f(x + tz) z$$

  - This implies:
    1. $\forall y \in X : (H_f(y)$ pos. semi-definite$) \Leftrightarrow f$ convex (proof in script)
    2. $\forall y \in X : (H_f(y)$ pos. definite$) \Rightarrow f$ strictly convex
  - Intuition: definiteness of the symmetric Hessian $\widehat{=}$ sign

- A lot of notation and definitions...
- Key take-aways:
  1. Gradients and Jacobians are the derivatives of multivariate functions
     - ...if the components (partial derivatives) are continuous; i.e. almost always
     - Intuition: summary of variation in fundamental directions of domain
  2. Taylor approximations give "good" polynomial approximations "close to" the approximation point
  3. Second derivatives of real-valued multivariate functions ("Hessian") can be obtained from differentiating the (transposed) gradient
  4. The definiteness of the Hessian determines convexity/concavity

- f is the instantaneous change of its accumulation

$\rightarrow$ If the integral measures accumulation, the function itself should be the integral's derivative!

- Idea: obtain integral operator by inverting the derivative operator

$$\frac{d}{dx} : D^1(X) \mapsto F_X, f \mapsto \frac{df}{dx}$$

- Issue: recall that inversion requires injectivity ("one-to-one")
  - $f(x) = 2x + 3$ vs. $f(x) = 2x$
  - Problem: constants cancel out when taking the derivative
  - Derivative is unique up to the constant!

Definition (Infimum and Supremum of a Set)

Let $X \subseteq \mathbb{R}$*. Then, the infimum* $\inf(X)$ *of* X *is the largest value smaller than any element of* X*, i.e.* $\inf(X) = \max\{a \in \mathbb{R} : \forall x \in X : x \geq a\}$*, and the supremum* $\sup(X)$ *of* X *is the smallest value larger than any element of* X*, i.e.* $\sup(X) = \min\{b \in \mathbb{R} : \forall x \in X : x \leq b\}$*.*

$\Rightarrow$ Generalized Maximum/Minimum

- Restrict attention to univariate, real-valued $f : X \mapsto \mathbb{R}$
- We can't invert $\frac{d}{dx}$, let's do the next best thing:

$$\int : F_X \mapsto \mathcal{P}(D^1(X)), f \mapsto \{\tilde{F} : X \mapsto \mathbb{R} : \frac{d\tilde{F}}{dx} = f\}$$

- Correspondence: set-valued mapping, **not** a function!
- We write $\int f = \{\tilde{F} : X \mapsto \mathbb{R} : \frac{d\tilde{F}}{dx} = f\}$ (pre-image of f under $\frac{d}{dx}$)
- Any $\tilde{F} \in \int f$ has the form $\tilde{F}(x) = F(x) + C$ for a $C \in \mathbb{R}$
  - F has no constant, i.e. $F(\min X) = 0$ or $\lim_{x \to \inf X} F(x) = 0$
  - F: accumulation at the left tail of the domain
  - Notation: $\tilde{F}(x) = \int f(x)dx = F(x) + C$

Theorem (Rules for Indefinite Integrals)

Let f, g be two integrable functions and let $a, b \in \mathbb{R}$ be constants, $n \in \mathbb{N}$. Then

- $\int (af(x) + g(x))dx = a \int f(x)dx + \int g(x)dx$,
- $\int x^n dx = \frac{x^{n+1}}{n+1} + C$ *if* $n \neq -1$ *and* $\int \frac{1}{x} dx = \ln(x) + C$,
- $\int e^x dx = e^x + C$ *and* $\int e^{f(x)} f'(x) dx = e^{f(x)} + C$,
- $\int (f(x))^n f'(x) dx = \frac{1}{n+1} (f(x))^{n+1} + C$ *if* $n \neq -1$ *and* $\int \frac{f(x)}{f'(x)} dx = \ln(f(x)) + C$.

Theorem (Integration by parts)

Let u, v be two differentiable functions. Then,

$$\int u(x)v'(x)dx = u(x)v(x) - \int u'(x)v(x)dx.$$

- Accumulation is unique up to initial level C: For any $\tilde{F} = F + C \in \int f$ and any $x, y \in X$: $\tilde{F}(y) - \tilde{F}(x) = F(y) - F(x)$

$\rightarrow$ Uniquely defined Definite Integral:

$$\int_x^y f(t)dt = \tilde{F}(y) - \tilde{F}(x), \quad \text{where} \quad \tilde{F}(x) \in \frac{d}{dx}^{-1}[\{f\}]$$

- Zero initial accumulation function if X is an interval:

$$F(x) = \int_a^x f(t)dt \quad \text{where} \quad a = \inf X$$

Theorem (Fundamental Theorem of Calculus)

Let X be an interval in $\mathbb{R}$ with $a = \inf(X)$ and $f : X \mapsto \mathbb{R}$. Suppose that f is integrable, and define $F := X \mapsto \mathbb{R}, x \mapsto \int_a^x f(t)dt$. Then, F is differentiable, and

$$\forall x \in X : F'(x) = \frac{dF}{dx}(x) = f(x).$$

- Proof (see script) is stunningly easy relative to the theorem's importance!
- Take-away
  - Fix initial accumulation to define a unique integral
  - This definite integral is inversely related to the derivative

- We have formally discussed univariate integration
- As with derivatives: if the multivariate integral exists, we can reduce its computation to univariate integrals!
- No formal details, rather only the "how-to"

Theorem (Fubini's theorem)

Let X and Y be two intervals in $\mathbb{R}$, let $f : X \times Y \to \mathbb{R}$ and suppose that f is continuous. Then, for any $I = I_x \times I_y \subseteq X \times Y$ with intervals $I_x \subseteq X$ and $I_y \subseteq Y$,

$$\int_I f(x,y)d(x,y) = \int_{I_x}\left(\int_{I_y} f(x,y)dy\right)dx,$$

and all the integrals on the right-hand side are well-defined.

General Fubini: for continuous $f : X \mapsto R$, $X \subseteq \mathbb{R}^n$

$$\int_I f(x_1,\dots,x_n)d(x_1,\dots,x_n) = \int_{I_1}\left(\cdots\left(\int_{I_n} f(x_1,\dots,x_n)dx_n\right)\cdots\right)dx_1.$$

Useful Corollary of Fubini:

Corollary (Integration of Multiplicatively Separable Functions)

Let $X_f \in \mathbb{R}^n, X_b \in \mathbb{R}^m$, $f : X_f \to \mathbb{R}$, $g : X_b \to \mathbb{R}$ continuous functions. Then, for any intervals $A \subseteq X_f$, $B \subseteq X_g$,

$$\int_{A\times B} f(x)g(y)d(x,y) = \left(\int_A f(x)dx\right)\left(\int_B g(y)dy\right).$$