Lecture 8: Derivatives and Integrals in $\mathbb{R}^n$

Juan Dodyk

WashU

Plan

Derivatives

1. Partial derivatives
2. Gradient
3. Hessian

Multiple integrals

1. Fubini

Functions of Multiple Variables

We'll work with functions $f : \mathbb{R}^n \to \mathbb{R}$, or $f : D \to \mathbb{R}$ for some $D \subset \mathbb{R}^n$.

For example, we want to minimize the mean squared error (MSE) of a statistical model given a sample.

This is a function of the vector of parameters of the model $\boldsymbol{\beta}$, say

$$f(\boldsymbol{\beta}) = \frac{1}{n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\beta} \cdot \boldsymbol{X}_{i\bullet})^2.$$

We can see a function with two variables, e.g., $f(x, y) = x^2 + 2y$, as a function that takes as input the vector $(x, y) \in \mathbb{R}^2$, i.e., a function $f : \mathbb{R}^2 \to \mathbb{R}$.

We know that taking derivatives is useful for minimizing/maximizing functions of one variable. It's also going to be useful for minimizing/maximizing functions of many variables.

Partial Derivatives

We want to determine how the value of a function changes when only one variable changes.

So, we take the derivative with respect to that variable, say x_i, keeping all other variables constant. We call that the **partial derivative** of f with respect to x_i, denoted by $\frac{\partial f}{\partial x_i}$.

Example. If $f(x_1, x_2) = x_1^2 + x_1 x_2$ then $\frac{\partial f}{\partial x_1} = 2x_1 + x_2$.

Same thing. If $f(x, y) = x^2 + xy$ then $\frac{\partial f}{\partial x} = 2x + y$.

We can think of the partial derivative $\frac{\partial y}{\partial x}$ as the “effect” on y of increasing x by one unit, keeping everything else constant.

Exercise

Question 1

Calculate

- $\dfrac{\partial}{\partial y}(x^2 + 2xy)$
- $\dfrac{\partial}{\partial x} e^{-x^2+y}$
- $\dfrac{\partial}{\partial z} e^{x-y+2z}$
- $\dfrac{\partial}{\partial \beta_i}(\boldsymbol{\beta} \cdot \boldsymbol{x})$, where $\boldsymbol{x}, \boldsymbol{\beta} \in \mathbb{R}^k$.

Gradient

The **gradient** ∇f is the vector of partial derivatives.

Example. If $f(x_1, x_2) = x_1^2 + x_1 x_2$ then $\nabla f(x_1, x_2) = (2x_1 + x_2, x_1)$.

The gradient is the direction in which the function grows the most (locally).

> The reason is that $f(x) \approx f(x_0) + \nabla f(x_0) \cdot (x - x_0)$, and
>
> $$\nabla f(x_0) \cdot (x - x_0) \leqslant \|\nabla f(x_0)\| \|x - x_0\|$$
>
> by Cauchy-Schwarz, attained for $x - x_0 = \nabla f(x_0)$.

So, if you want to maximize f, you can take some x, and then go to $x + \alpha \nabla f(x)$ for some $\alpha > 0$, and keep doing this. If you want to minimize f, go to $x - \alpha \nabla f(x)$. This is **gradient descent**, this is how deep learning models are trained.

Exercise

Question 2

Calculate the gradients of the following functions:

- $f(x, y) = x^2 + 2xy$,
- $f(x, y) = e^{-x^2+y}$,
- $f(x, y, z) = e^{x-y+2z}$,
- $f(\boldsymbol{\beta}) = \boldsymbol{x} \cdot \boldsymbol{\beta}$.

Optimization

If $f : D \to \mathbb{R}$ with $D \subset \mathbb{R}^n$, $\boldsymbol{x}_0$ is in the interior of D, and f is maximized at $\boldsymbol{x}_0$, then

$$\nabla f(\boldsymbol{x}_0) = 0.$$

Why? Every partial derivative should be zero, because otherwise we could increase f by increasing/decreasing one variable (keeping the rest fixed).

Second-order Derivatives

If f is a function of $\boldsymbol{x}$, the partial derivative $\frac{\partial f}{\partial x_i}$ is also a function of $\boldsymbol{x}$. We can take another partial derivative, e.g., $\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}$. We write $\boldsymbol{\frac{\partial^2 f}{\partial x_j \partial x_i}}$.

If the second-order partial derivatives are continuous then $\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}$, i.e., it doesn't matter in what order we take the partial derivatives. We write $\boldsymbol{\frac{\partial^2 f}{\partial^2 x_i}} = \frac{\partial^2 f}{\partial x_i \partial x_i}$.

The **Hessian** $\nabla^2 f$ is the $n \times n$ matrix of second-order derivatives:

$$\boldsymbol{\nabla^2 f} = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial^2 x_n} \end{pmatrix}$$

Example

If $f(x, y) = x^2 + 2xy$, $\dfrac{\partial f}{\partial x} = 2x + y$, $\dfrac{\partial f}{\partial y} = 2x$, $\dfrac{\partial^2 f}{\partial^2 x} = 2$, $\dfrac{\partial^2 f}{\partial x \partial y} = 2$, and $\dfrac{\partial^2 f}{\partial^2 y} = 0$, so

$$\nabla^2 f = \begin{pmatrix} 2 & 2 \\ 2 & 0 \end{pmatrix}.$$

If $g(x, y, z) = e^{x-y+2z}$, $\dfrac{\partial g}{\partial x} = g$, $\dfrac{\partial g}{\partial y} = -g$, $\dfrac{\partial g}{\partial z} = 2g$, $\dfrac{\partial^2 g}{\partial^2 x} = g$, $\dfrac{\partial^2 g}{\partial x \partial y} = -g$, $\dfrac{\partial^2 g}{\partial x \partial z} = 2g$, $\dfrac{\partial^2 g}{\partial^2 y} = g$, $\dfrac{\partial^2 y}{\partial z \partial =} - 2g$, and $\dfrac{\partial^2 g}{\partial^2 z} = 4g$, so

$$\nabla^2 g = \begin{pmatrix} g & -g & 2g \\ -g & -g & -2g \\ 2g & -2g & 4g \end{pmatrix} = g \begin{pmatrix} 1 & -1 & 2 \\ -1 & -1 & -2 \\ 2 & -2 & 4 \end{pmatrix}.$$

Exercise

Question 3

Calculate $\nabla^2 f$ for $f(x, y) = e^{-(x-y)^2}$.

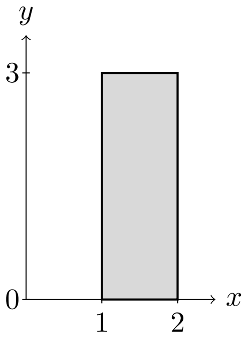
Notice. An alternative notation for the gradient and Hessian is $\boldsymbol{Df}$ and $\boldsymbol{D^2f}$.

The Hessian is the generalization of the second derivative $f''(x)$, and it's useful because it tells us if the function is concave or convex. We'll see this later.

Intervals in n Dimensions

The interval $[a, b] \times [c, d]$ is the set $\{(x, y) \in \mathbb{R}^2 : x \in [a, b] \text{ and } y \in [c, d]\}$.

Example. $[1, 2] \times [0, 3]$



In general, $[a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$. Also, $[a, b]^2 = [a, b] \times [a, b]$.

Integrals of Functions of Several Variables

If $f : D \to [0, +\infty)$ with $D \subset \mathbb{R}^2$ then the integral

$$\iint_D f(x, y)\, dxdy$$

is the volume of the region between the xy plane and the graph of the function f.

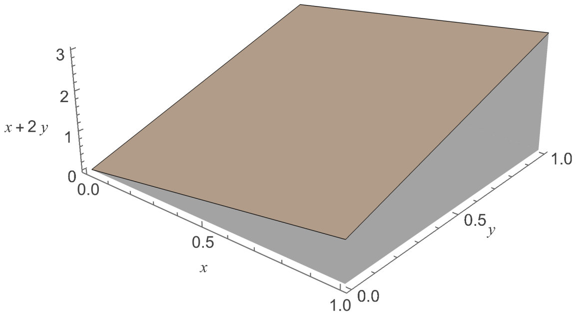
Fubini. (Special case.) If $D = [a, b] \times [c, d]$ then

$$\iint_D f(x, y)\, dxdy = \int_c^d \int_a^b f(x, y)\, dxdy = \int_a^b \int_c^d f(x, y)\, dydx.$$

In general, if $f : D \to [0, +\infty)$ with $D \subset \mathbb{R}^n$ the integral $\int_D f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$ is the “measure” of the region below the graph of f.

Example

$$\iint_{[0,1]^2} (x+2y)\,dxdy = \int_0^1 \int_0^1 (x+2y)\,dxdy = \int_0^1 \left[\frac{x^2}{2} + 2yx\right]_{x=0}^1 dy$$
$$= \int_0^1 \left(\frac{1}{2} + 2y\right) dy = \left[\frac{1}{2}y + y^2\right]_0^1 = \frac{3}{2}.$$



Exercise

Question 4

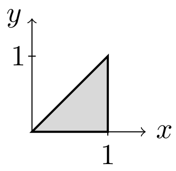
Calculate $\iint_{[0,1]\times[0,2]} e^{-x-y}\,dxdy$.

Question 5

Calculate $\iint_{[0,+\infty)^2} e^{-x-y}\,dxdy = \int_0^{+\infty}\int_0^{+\infty} e^{-x-y}\,dxdy$.

We can extend the definition to f negative in the same way we did for functions of one variable.

We can use Fubini over domains that are not intervals. For example, consider this triangle:



We can see it as the set of points (x, y) with $0 \leqslant x \leqslant 1$ and $0 \leqslant y \leqslant x$. So, if

$$D = \{(x, y) \in \mathbb{R}^2 : 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant x\},$$

we should have

$$\iint_D f(x, y)\, dxdy = \int_0^1 \int_0^x f(x, y)\, dydx = \int_0^1 \int_y^1 f(x, y)\, dxdy.$$