CSE 840: Computational Foundations of Artificial Intelligence November 01, 2023

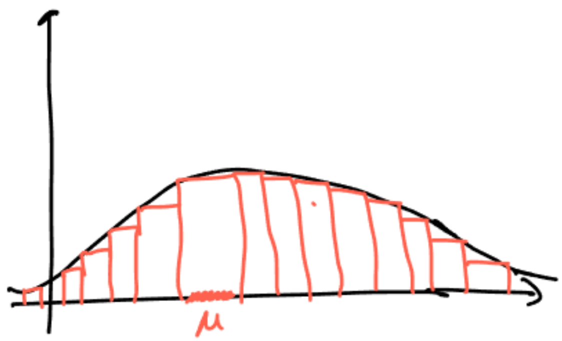
Lebesgue Integral on R^n, Differentiation on R^n

Instructor: Vishnu Boddeti *Scribe: Samia Islam, Mk Bashar, Patrick Ancel*

1 The Lebesgue Integral on $\mathbb{R}^n$

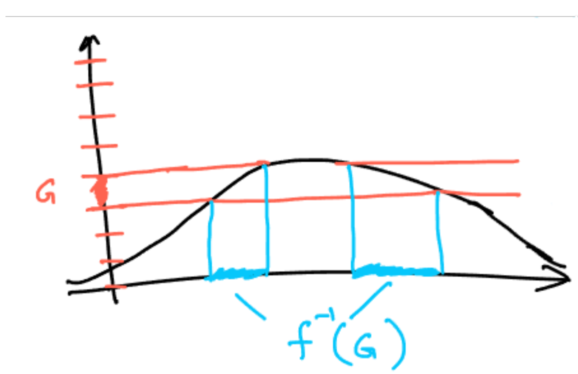
Intuition: Riemann Integral:

- bounded
- continuous
- finite set of rectangles



Lebesgue Integral:

- not bounded
- need not be continuous
- countable sets



Definition 1 *A function* $f : (\Omega_1, \mathbb{A}_1) \to (\Omega_2, \mathbb{A}_2)$ *between two measurable spaces is called* <u>*measurable*</u> *if pre-image of any measurable set is measurable:*

$$\forall A_2 \in \mathbb{A} : f^{-1}(A_2) \in \mathbb{A}_1 \quad where,\ f^{-1}(A_2) := \{x \in \Omega_1 | f(x) \in A_2\}$$

$(\Omega, \mathbb{A}), (\mathbb{R}, \mathbb{B}(\mathbb{R}))$

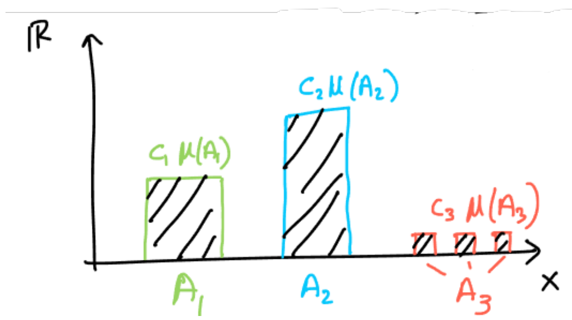
Characteristic function (also indicator function)

$$\chi_A : \Omega \to \mathbb{R}, \chi_A(\omega) := \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

Definition 2 $\phi : \mathbb{R}^n \to \mathbb{R}$ *is called a simple funciton if there exist measurable sets* $A_i \subset \mathbb{R}^n$, $c_i \in \mathbb{R}$ *such that,*

$$\phi(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x)$$

$$I = c_1\mu(A_1) + c_2\mu(A_2) + c_3\mu(A_3)$$



$$\phi(x) = c_1\chi_{A_1}(x) + c_2\chi_{A_2}(x) + c_3\chi_{A_3}(x)$$

$$I(\phi) = \int \phi d\mu = \sum_{i=1}^{n} c_i\mu(A_i) \to \text{Lebesgue integral for simple function}$$

Problem : $3 \cdot \infty - 2 \cdot \infty$??

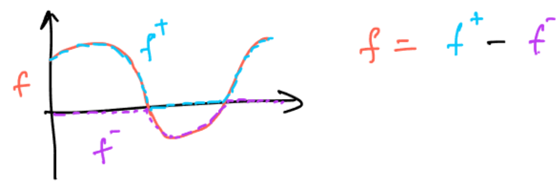
For a function $f^+ : \mathbb{R}^n \to [0, \infty)$ we define its Lebesgue integral as,

$$\int f^+ d\mu = \sup \left\{ \int \phi d\mu | \phi \leq f, \phi \text{ is simple} \right\}$$

(might be ∞)

For a general function $f : \mathbb{R}^n \to \mathbb{R}$ we split the function into positive and negative parts:

$$f = f^+ - f^-, \quad f^+ \geq 0, f^- \geq 0 \quad \text{where, } f^+ = \begin{cases} f(x), & f(x) \geq 0 \\ 0 & otherwise \end{cases}$$



Note: f^+, f^- are measurable if f is measurable. If both f^+ and f^- satisfy $\int f^+ d\mu < \infty$ and $\int f^- d\mu < \infty$, then we call f integrable and define

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

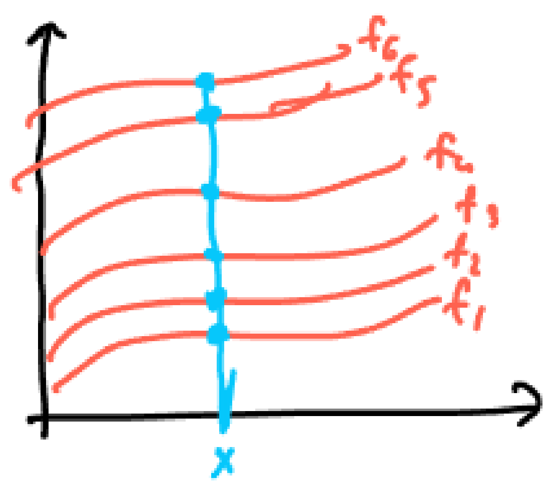
Much more powerful notion than Reimann Integral.

Example: $\int \chi_{\mathbb{Q}} d\mu = 1 \cdot \mu(\mathbb{Q}) = 0$

2 Two Important Theorems

Theorem 3 ***Monotone Convergence:*** *Consider a sequence of functions $f_n : \mathbb{R}^n \to [0, \infty)$ that is pointwise non-decreasing: $\forall x \in \mathbb{R}^n, f_{k+1}(x) \geq f_k(x)$. Assume that all f_k are measurable, and that the pointwise limit exists $\forall x : lim f_k(x) := f(x)$. Then,*

$$\int \lim_{k \to \infty} f_k(x) dx = lim_{k \to \infty} \int f_k(x) dx$$



Theorem 4 ***Dominated Convergence:*** *$f_k : \mathbb{B} \to \mathbb{R}, |f_k(x)| \leq g(x)$ on $\mathbb{B}, g(x)$ is integrable. Assume that the pointwise limit exists: $\forall x \in \mathbb{B}, f(x) := \lim_{n \to \infty} f_n(x)$. Then,*

$$\int \lim_{k \to \infty} f_k(x) dx = \lim_{k \to \infty} \int f_k(x) dx$$

3 Partial Derivatives on $\mathbb{R}^n$

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$

Definition 5 *The function f is called partially differentiable with respect to variable x_j at point $\xi \in \mathbb{R}^n$ if the function*

$$g : \mathbb{R} \to \mathbb{R}$$

$$x_j \mapsto g(x_j) := f(\xi_1, \xi_2, \ldots, \xi_{j-1}, x_j, \xi_{j+1}, \ldots, \xi_n)$$

is differentiable at $\xi_j \in \mathbb{R}$. All variables aside from x_j are treated as constants, making g a function of one variable.

The notation for the partial derivative is:

$$\frac{\partial f(\xi)}{\partial x_j} = \lim_{h \to 0} \frac{f(\xi + \mathbf{e}_j h) - f(\xi)}{h}$$

Here, h is a scalar, and $\mathbf{e}_j$ is the j-th unit vector, which has a 1 at the j-th index and zeros everywhere else.

Example: Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ where $f(x) = x_1^2 + x_2^2 x_1$. The derivative with respect to x_1 is computed by treating x_2 as a constant, so $\frac{\partial f}{\partial x_1} = 2x_1 + x_2^2$.

If all partial derivatives exist, then the vector of all partial derivatives is called the gradient.

$$\operatorname{grad}(f)(\xi) = \nabla f(\xi) = \begin{bmatrix} \frac{\partial f(\xi)}{\partial x_1} \\ \vdots \\ \frac{\partial f(\xi)}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n$$

If $f : \mathbb{R}^n \to \mathbb{R}^m$, we can decompose f into its m component functions.

$$f = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}$$

We define the Jacobian matrix:

$$Df(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} —(\nabla f_1(x))^T— \\ \vdots \\ —(\nabla f_m(x))^T— \end{bmatrix} \in \mathbb{R}^{m \times n}$$

The i-th row of the Jacobian matrix is the gradient of f_i.

Caution: Even if all partial derivatives exist at ξ, we do not know if f is continuous at ξ.

Example: Consider $f : \mathbb{R}^2 \to \mathbb{R}$.

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } x = y = 0 \end{cases}$$

For $(x, y) \neq (0,0)$,

$$\nabla f(x,y) = \left(y\frac{y^2 - x^2}{(x^2+y^2)^2}, x\frac{x^2 - y^2}{(x^2+y^2)^2} \right)$$

$\nabla f(0,0) = 0$ since for all x, $f(x,0) = 0$, and for all y, $f(0,y) = 0$. But f is not continuous at $(0,0)$.

4 Total Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $\xi \in U$. The function f is differentiable at ξ if there exists a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ such that for $h \in \mathbb{R}^n$,

$$f(\xi + h) - f(\xi) = L(h) + r(h)$$

$$\text{with } \lim_{h \to 0} \frac{r(h)}{|h|} = 0$$

That is, f is differentiable at ξ if it can be approximated locally by a linear mapping. The difference between $f(\xi + h)$ and $f(\xi)$ is a linear mapping plus residue $r(h)$. The residue goes to zero as h goes to zero.

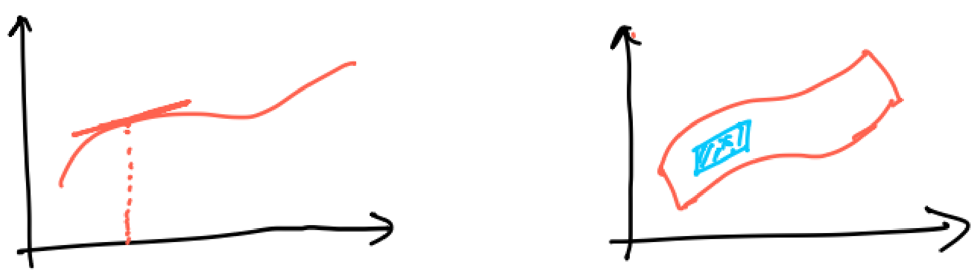


Figure 1: A linear mapping approximating a function in 2D (left) and in 3D (right).

Intuition: f is "locally linear"

Theorem 6 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at ξ. Then f is continuous at ξ, and the linear functional L coincides with the gradient:* $f(\xi+h) - f(\xi) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\xi) \cdot h_j + r(h) = \langle \nabla f(\xi), h \rangle + r(h)$

If $f : \mathbb{R}^n \to \mathbb{R}^m$, it is differentiable if all coordinate functions $f_1, f_2, \ldots, f_m$ are differentiable. Then all partial derivatives exist and $L(h) = (\text{Jacobian matrix}) \cdot h$.

Theorem 7 *If all partial derivatives exist and are all continuous, then f is differentiable.*

Warning: If partial derivatives exist, but are not continuous, then f does not need to be differentiable.

5 Directional Derivatives

The idea is to compute derivatives along an arbitrary direction, not just the coordinate axes.

Definition 8 *Assume $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and $\mathbf{v} \in \mathbb{R}^n$ with $\|\mathbf{v}\| = 1$. The* *directional derivative* *of f at ξ in the direction of $\mathbf{v}$ is defined as,*

$$D_{\mathbf{v}} f(\xi) = \lim_{t \to 0} \frac{f(\xi + t\mathbf{v}) - f(\xi)}{t}$$

In this equation, $t \in \mathbb{R}$ is a scalar and $\mathbf{v} \in \mathbb{R}^n$ is a unit vector corresponding to a direction.

Theorem 9 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at ξ. Then all the directional derivatives exist, and we can compute them as,*

$$D_{\mathbf{v}} f(\xi) = (\nabla f(\xi))^T \mathbf{v} = \sum_{i=1}^{n} v_i \frac{\partial f}{\partial x_i}(\xi)$$

In this equation, $v_i \in \mathbb{R}$ is a scalar, and $\mathbf{v}$ is a vector.

The partial derivative in any direction can be expressed as a weighted linear combination of derivatives along the axes.

The largest value of all directional derivatives is attained in the direction of the gradient: $\mathbf{v} = \frac{\nabla f(\xi)}{\|\nabla f(\xi)\|}$