Module 10: Differentiation and Integration

Operational math bootcamp

Emma Kroell

University of Toronto

July 28, 2023

Outline

- Differentiation on $\mathbb{R}$
  - Mean value theorem
  - l'Hôpital's rule
  - Smoothness classes
- Integration on $\mathbb{R}$
  - Riemann sums and Riemann integral
  - Integration rules
  - Drawbacks of Riemann integration

Block matrices

Definition

A block matrix is a matrix that can be broken into sections called blocks, which are smaller matrices.

Example

$$\begin{bmatrix} 2 & 1 & 0 & 1 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

Definition

A square matrix is called *block diagonal* if it can be written as a block matrix where the main-diagonal blocks are all square matrices and the off-diagonal blocks are all zero.

Example

The matrix

$$\begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

is block diagonal.

Definition

A vector $\mathbf{v}$ is called a *generalized eigenvector* of A corresponding to an eigenvalue λ if there exists $k \geq 1$ such that

$$(A - \lambda I)^k \mathbf{v} = 0.$$

The set of generalized eigenvectors of an eigenvalue λ (plus $\mathbf{0}$) is called the *generalized eigenspace* of λ.

Proposition

The algebraic multiplicity of an eigenvalue λ is the same as the dimension of the corresponding generalized eigenspace.

Theorem (Jordan decomposition theorem)

For any operator A there exists a basis such that A is block diagonal with blocks that have eigenvalues on the diagonal and 1s on the upper off-diagonal. In other words, A can be written in the form

$$A = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_k \end{bmatrix}$$

where the blocks J_i on the main diagonal are Jordan block of the form

$$\begin{bmatrix} \lambda \end{bmatrix}, \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}, \textit{ etc.}$$

This form is called Jordan canonical form.

Connection to algebraic and geometric multiplicity:

- The algebraic multiplicity of an eigenvalue λ is the number of times λ appears on the diagonal.
- The geometric multiplicity of λ is the number of Jordan blocks associated with λ.

Why is Jordan form useful?

Singular value decomposition

- $A^T A$ is symmetric
- therefore it is orthogonally diagonalizable and has real eigenvalues
- In fact, the eigenvalues are non-negative (exercise)

Definition

Let A be an $m \times n$ matrix. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A^T A$. Then the *singular values* of A are defined as

$$\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_n = \sqrt{\lambda_n}.$$

Theorem (Singular value decomposition)

If A is an $m \times n$ matrix of rank k, then we can write

$$A = U\Sigma V^T$$

where Σ is an $m \times n$ matrix of the form

$$\begin{bmatrix} D_{k\times k} & 0_{k\times(n-k)} \\ 0_{(m-k)\times k} & 0_{(m-k)\times(n-k)} \end{bmatrix},$$

D is a diagonal matrix with the singular values of A, $\sigma_1, \ldots, \sigma_k$, on the diagonal and U and V are both orthogonal matrices (of size $m \times m$ and $n \times n$, respectively).

Uses of SVD:

Differences between JCF and SVD:

LU-decomposition

Definition

The LU-decomposition of a square matrix A is the factorization of A into a lower triangular matrix L and an upper triangular matrix U as follows:

$$A = LU.$$

Why is this useful? Consider the linear system $A\mathbf{x} = \mathbf{b}$

Recall: orthonormal basis

Definition

Two vectors $\mathbf{x}, \mathbf{y} \in V$ are called *orthogonal* if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, denoted $\mathbf{x} \perp \mathbf{y}$. We call them *orthonormal* if additionally the vectors are normalized, i.e. $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$. A basis $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of V is called *orthonormal basis (ONB)*, if the vectors are pairwise orthogonal and normalized.

Starting from a set of linearly independent vectors, we can construct another set of vectors which are orthonormal and span the same space using the **Gram-Schmidt Algorithm**.

QR-decomposition

Definition (QR-decomposition)

The QR-decomposition of an $m \times n$ matrix A with linearly independent column vectors is the factorization of A as follows:

$$A = QR,$$

where Q is an $m \times n$ matrix with orthonormal column vectors and R is an $n \times n$ invertible upper triangular matrix.

One obtains the factorization by applying the Gram-Schmidt algorithm to the columns of A. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be the column vectors of A. Let $\mathbf{q}_1, \ldots, \mathbf{q}_n$ be the orthonormal vectors obtained by applying Gram Schmidt. Then one can write:

$$
\begin{aligned}
\mathbf{u}_1 &= \langle \mathbf{u}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n \\
\mathbf{u}_2 &= \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n \\
&\vdots \\
\mathbf{u}_n &= \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n
\end{aligned}
$$

Thus the orthonormal vectors obtained using Gram-Schmidt form the columns of Q, while R is the terms needed to go between the columns of A and thsoe of Q, i.e.

$$
R = \begin{bmatrix}
\langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \ldots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \\
0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \ldots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle
\end{bmatrix}.
$$

Why use QR-decomposition?

Differentiation

Derivative

Recall the definition of the derivative:

Definition

A function $f : (a, b) \to \mathbb{R}$ is *differentiable* at $x \in (a, b)$ if

$$L := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

exists. L is the *derivative* of f at x, denoted $L = f'(x)$. If f is differentiable at every $x \in (a, b)$, we say f is *differentiable*.

Proposition

The following are key rules for differentiation:

1. If f is differentiable at x, then it is continuous at x.
2. The derivative of a constant function is zero.
3. If f and g are differentiable at x, then so is $f+g$ with $(f+g)'(x) = f'(x) + g'(x)$.
4. Product rule: If f and g are differentiable at x, then so is fg with
$$(fg)'(x) = f'(x)g(x) + f(x)g'(x).$$
5. Quotient rule: If f and g are differentiable at x and $g(x) \neq 0$, then so is f/g with
$$(f/g)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}.$$
6. Chain rule: If f is differentiable at x and g is differentiable at $f(x)$, then so is $g \circ f$ with
$$(g \circ f)'(x) = g'(f(x))f'(x).$$

Lemma

If $f : (a, b) \to \mathbb{R}$ is differentiable on (a, b) and achieves a (local) maximum or (local) minimum at $c \in (a, b)$, then $f'(c) = 0$.

Proof

Proof continued

Next we recall a theorem we proved in greater generality when we discussed compactness in the topology section:

Theorem (Extreme value theorem)

If $f : [a, b] \to \mathbb{R}$ is continuous, then f attains a maximum and a minimum, i.e. there exists $c, d \in [a, b]$ such that

$$f(c) \leq f(x) \leq f(d) \quad \forall x \in [a, b].$$

This theorem is used to prove the following important result:

Theorem (Mean value theorem)

If f is continuous on $[a, b]$ and differentiable on (a, b), then there exists $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a).$$

Proof.

$\square$

Corollary

If $f : (a, b) \to \mathbb{R}$ is differentiable and has a bounded derivative (i.e. $|f'(x)| \leq M$ for some $M > 0$ and for all $x \in (a, b)$), then f is Lipschitz.

Proof.

$\square$

l'Hôpital's rule

Theorem (l'Hôpital's rule)

If f, g are differentiable on (a, b), where a, b may be $\pm\infty$, and $\lim_{x\to b} f(x) = 0 = \lim_{x\to b} g(x)$, or both limits equal $\pm\infty$, then

$$\lim_{x\to b} \frac{f'(x)}{g'(x)} = L$$

implies

$$\lim_{x\to b} \frac{f(x)}{g(x)} = L$$

Example

$$\lim_{x \to 0} \frac{5^x - 2^x}{x^2 - x}$$

$$\lim_{x \to -\infty} xe^x$$

Higher order derivatives

Definition

We define higher-order derivatives inductively as $f^{(r)}(x) = (f^{(r-1)})'(x)$. If $f^{(r)}$ exists (at x), we say that f is r^{th}-order differentiable (at x).

Definition

If $f^{(r)}$ exists for all $r \in \mathbb{N}$ and for all $x \in (a, b)$, then we say f is infinitely differentiable or *smooth*. We denote this $f \in C^{\infty}$.

Smoothness classes

Definition

If f is differentiable and its derivative $f'(x)$ is continuous, we say that f is *continuously differentiable*, and that $f \in C^1$. If $f^{(r)}$ exists and is continuous, we say that $f \in C^r$. If f is continuous, we say $f \in C^0$.

Since differentiability implies continuity, we have $C^\infty \subset \cdots \subset C^2 \subset C^1 \subset C^0$.

Example

- The function $f(x) = |x|$ is C^0 but not C^1.
- The function $f(x) = x|x|$ is C^1 but not C^2.
- $f(x) = e^x$ and $f(x) = x$ are smooth functions, i.e. in C^∞.

Integration

Riemann integration

Definition (Riemann sum)

Let $f : [a, b] \to \mathbb{R}$ be a function. We call a set of points $P = \{x_0, \ldots, x_n\} \subseteq [a, b]$ a *partition* of $[a, b]$ if the following holds

$$a = x_0 \leq x_1 \leq \ldots \leq x_{n-1} \leq x_n = b.$$

We call the largest sub-interval of the partition P the *mesh* of P, denoted $|P|$, i.e.

$$|P| = \max_{i=1,\ldots,n} (x_i - x_{i-1}).$$

Definition continued (Riemann sum)

Given a partition $P = \{x_0, \ldots, x_n\} \subseteq [a, b]$ of $[a, b]$ and a set of points $T = \{t_1, \ldots, t_n\} \subseteq [a, b]$ such that $t_i \in [x_{i-1}, x_i]$ for $i = 1, \ldots, n$, we define the *Riemann sum* $R(f, P, T)$ corresponding to f, P, T as

$$R(f, P, T) = \sum_{i=1}^{n} f(t_i)(x_i - x_{i-1}) := \sum_{i=1}^{n} f(t_i)\Delta x_i,$$

where we used $\Delta x_i = x_i - x_{i-1}$.

The idea is to define the Riemann integral as the "limit" of the Riemann sums over all partition such that their mesh is becoming arbitrarily small:

Definition (Riemann integrable)

A function $f : [a, b] \to \mathbb{R}$ is called *Riemann integrable* if there exists $I \in \mathbb{R}$ such that for all $\epsilon > 0$, there exists a $\delta > 0$ such that for any partition $P = \{x_0, \dots, x_n\}$ of $[a, b]$ with $|P| < \delta$ and set of points $T = \{t_1, \dots, t_n\} \subseteq [a, b]$ such that $t_i \in [x_{i-1}, x_i]$ for $i = 1, \dots, n$ we have $|R(f, P, T) - I| < \epsilon$.
We say that I is the Riemann integral of f, denoted $I = \int_a^b f(x)\mathrm{d}x$.

If f is Riemann integrable, then I is unique.

Let $\mathcal{R}([a,b])$ denote the set of functions that are Riemann integrable on $[a,b]$.

Theorem

Riemann integration is linear, i.e. if $f, g \in \mathcal{R}([a,b])$ and $c \in \mathbb{R}$, then $f + cg \in \mathcal{R}([a,b])$.

Proof

Proof continued

Proposition (Rules for integration on $[a, b]$)

1. The constant function $f(x) = c$ is integrable and its integral is $c(b-a)$.
2. If f is Riemann integrable, then it is bounded.
3. If $f, g \in \mathcal{R}([a, b])$ and $f(x) \leq g(x)$ for all $x \in [a, b]$, then
$$\int_a^b f(x)dx \leq \int_a^b g(x)dx.$$
4. If $f \in \mathcal{R}([a, b])$ and $g : [c, d] \to [a, b]$ is a continuously differentiable bijection with $g' > 0$, then
$$\int_a^b f(y)dy = \int_c^d f(g(x))g'(x)dx.$$
5. If $f, g : [a, b] \to \mathbb{R}$ are differentiable and $f', g' \in \mathcal{R}([a, b])$, then
$$\int_a^b f(x)g'(x)dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)dx.$$

Theorem (Fundamental Theorem of Calculus)

First part:

If $f : [a, b] \to \mathbb{R}$ is Riemann integrable then its indefinite integral

$$F(x) = \int_a^x f(t)dt$$

is a continuous function of x. In addition, the derivative of F exists and $F'(x) = f(x)$ at all $x \in [a, b]$ where f is continuous.

Second part:

Let $f : [a, b] \to \mathbb{R}$ and let F be a continuous function on $[a, b]$ with antiderivative f on (a, b), i.e. $F'(x) = f(x)$. Then if F is Riemann integrable on $[a, b]$,

$$\int_a^b f(x)dx = F(b) - F(a).$$

Drawbacks of the Riemann integral

- Riemann integration has many nice properies, but it also has some drawbacks
- To see this, we first introduce a nice alternative characterization of Riemann integrability
- Instead of looking at all the Riemann sums, we can restrict our attention to two special forms of the sum

Definition

Given a function $f\colon [a,b] \to \mathbb{R}$ and a partition $P = \{x_0, \ldots, x_n\}$ of $[a,b]$, we define the *lower* and *upper sum* of f via

$$L(f,P) = \sum_{i=1}^{n} m_i \Delta x_i, \qquad U(f,P) = \sum_{i=1}^{n} M_i \Delta x_i,$$

where $m_i = \inf\{f(t)\colon t \in [x_{i-1}, x_i]\}$ and $M_i = \sup\{f(t)\colon t \in [x_{i-1}, x_i]\}$. We define the *lower* and *upper integral* of f to be

$$\underline{I} = \sup_P L(f,P), \qquad \overline{I} = \inf_P U(f,P).$$

Since f is defined on a compact set, it will be bounded and hence the lower and upper integral exist and are finite.

One can think of the lower and upper integral as lower and upper bounds for the Riemann integral.

Theorem

Let $f : [a, b] \to \mathbb{R}$ be a function. Then f is Riemann integrable if and only if $\underline{I} = \overline{I}$ and we have $\underline{I} = \overline{I} = I$.

A function that is not Riemann integrable

$$f\colon [0,1] \to \mathbb{R}\colon x \mapsto \begin{cases} 0 & \text{if } x \in \mathbb{Q}, \\ 1 & \text{otherwise.} \end{cases}$$

Is this function Riemann integrable? Should it be integrable?

The End

References

Charles C. Pugh (2015). *Real Mathematical Analysis*. Undergraduate Texts in Mathematics. https://link-springer-com.myaccess.library.utoronto.ca/book/10.1007/978-3-319-17771-7