Optimization

Shuoqi Sun

Tepper School of Business, CMU

Plan

- Review of relevant concepts
- First order approach to optimization(Lagrangian method, KKT conditions...
- Comparative static analysis
- Calculus of Variations

Convex and Concave Functions

- Let $X \subseteq \mathbb{R}^n$ be convex and $f : X \to \mathbb{R}$.
- f is said to be **convex** if:

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

- f is said to be **strictly convex** if:

$$\forall x, y \in X, \forall \lambda \in (0,1) \;\; f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

- f is said to be **(strictly) concave** if $-f$ is (strictly) convex.
- f is said to be concave if:

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

- f is said to be strictly concave if:

$$\forall x, y \in X, \forall \lambda \in (0,1) \;\; f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda) f(y)$$

Affine Transformations

- $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine if $f(x) = Ax + b$ for some $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.
- The image of a convex set under an affine function is convex.
- The inverse image of a convex set under an affine function is convex.
- Affine functions are both convex and concave.

Fundemental properties of convex functions

- (Chordal Slope Lemma)Let $(a, b) \subset \mathbb{R}$ and suppose $f : (a, b) \to \mathbb{R}$ is convex. Then for any $x, y, z \in (a, b)$ with $x < y < z$,

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(x)}{z - x} \leq \frac{f(z) - f(y)}{z - y}$$

- Any convex functions defined on a open interval is Lipschitiz continuous.
- A convex function is differentiable anywhere except at finitely many points.

A quick refresher on differentiation

First-Order Condition

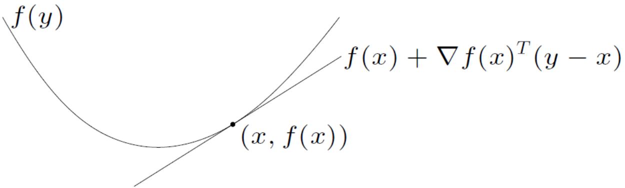
- Let $X \subseteq \mathbb{R}^n$ be convex and $f : X \to \mathbb{R}$ be differentiable.
- **Gradient** of f:

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

- f is convex if and only if:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \ \forall x, y \in X$$

- First-order approximation of f is a global underestimator.



Second-Order Condition

- Let $X \subseteq \mathbb{R}^n$ be convex and $f : X \to \mathbb{R}$ be twice differentiable.
- **Hessian** of f is an $n \times n$ matrix of second derivatives:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n x_1} & \frac{\partial^2 f(x)}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

- f is convex (concave) if and only if $\nabla^2 f(x)$ is positive (negative) semi-definite for all $x \in X$.
- If $\nabla^2 f(x)$ is positive (negative) definite for all $x \in X$, f is strictly convex (concave).

Example

- $f(x_1, x_2) = x_1^2 + 2x_1x_2 + x_2^2$ is a convex function on $\mathbb{R}^2$.

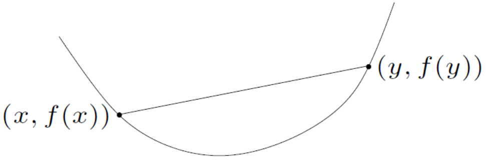
$$H(x_1, x_2) = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

- $f(x_1, x_2) = -x_1^2 - x_1x_2 - 2x_2^2$ is a strictly concave function on $\mathbb{R}^2$.

$$H(x_1, x_2) = \begin{bmatrix} -2 & -1 \\ -1 & -4 \end{bmatrix}$$

Convex Functions: Example

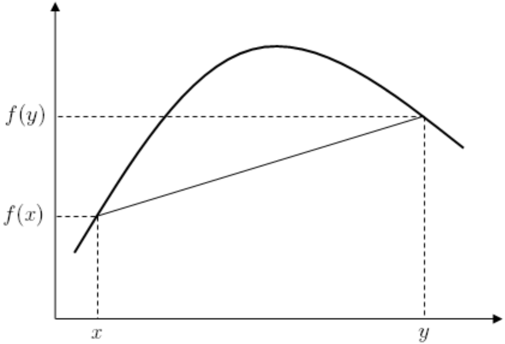
- Affine functions: $f(x) = a^T x + b$
- Exponential: e^{ax}, for any $a \in \mathbb{R}$
- Powers: x^α on $\mathbb{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- Powers of absolute value: $|x|^p$ on $\mathbb{R}$, for $p \geq 1$
- Negative entropy: $x \log x$ on $\mathbb{R}_{++}$
- Norms: $||x||_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$ for $p \geq 1$; $||x||_\infty = \max_k |x_k|$



- Sums of convex functions are convex.

Concave Functions: Example

- Affine functions: $f(x) = a^T x + b$
- Logarithm: $\log x$ on $\mathbb{R}_{++}$
- Powers: x^α on $\mathbb{R}_{++}$, for $0 \leq \alpha \leq 1$



- Sums of concave functions are concave.

Quasi-Convex and Quasi-Concave Functions

- Let $X \subseteq \mathbb{R}^n$ be convex and $f : X \to \mathbb{R}$.
- f is said to be **quasi-convex** if:

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

- f is said to be **strictly quasi-convex** if:

$$\forall x, y \in X, \forall \lambda \in (0, 1) \;\; f(\lambda x + (1 - \lambda)y) < \max\{f(x), f(y)\}$$

- f is said to be **(strictly) quasi-concave** if $-f$ is (strictly) quasi-convex.
- f is said to be quasi-concave if:

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

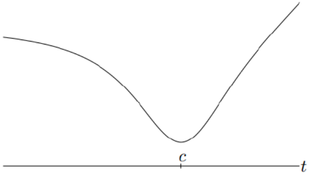
- f is said to be strictly quasi-concave if:

$$\forall x, y \in X, \forall \lambda \in (0, 1) \;\; f(\lambda x + (1 - \lambda)y) > \min\{f(x), f(y)\}$$

- f is **quasi-linear** if it is both quasi-convex and quasi-concave.

Quasi-Convex Functions on $\mathbb{R}$

- f is quasi-convex if and only if at least one of the following conditions holds:
  - f is non-decreasing.
  - f is non-increasing.
  - There exists a point $c \in X$ such that for $t \leq c$ (and $t \in X$), f is non-increasing, and for $t \geq c$ (and $t \in X$), f is non-decreasing.



- Any monotonically increasing (decreasing) f is quasi-linear.

Quasi-Convex and Quasi-Concave Functions: Example

- $\sqrt{|x|}$ is quasi-convex on $\mathbb{R}$
- $ceil(x) = \inf\{z \in \mathbb{Z} | z \geq x\}$ and $floor(x) = \sup\{z \in \mathbb{Z} | z \leq x\}$ are quasilinear
- $\log x$ is quasilinear on $\mathbb{R}_{++}$
- $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbb{R}^2_{++}$
- e^x is convex, quasi-concave, quasi-convex but not concave.
- Sums of quasi-convex functions are not necessarily quasi-convex.

Upper and Lower Contour Sets

- The upper contour set of $f : X \to \mathbb{R}$ for a scalar level $c \in \mathbb{R}$ is $U_f(c) = \{x \in X | f(x) \geq c\}$.
- The lower contour set of $f : X \to \mathbb{R}$ for a scalar level $c \in \mathbb{R}$ is $L_f(c) = \{x \in X | f(x) \leq c\}$.
- All lower (upper) contour sets of convex (concave) functions are convex.
- $f : X \to \mathbb{R}$ is quasi-convex (concave) if and only if all lower (upper) contour sets are convex.
- If f is quasilinear, then every level set $\{x | f(x) = c\}$ is convex.

Optimization Problem in Standard Form

- Let $X \subseteq \mathbb{R}^n$ be convex and $f : X \to \mathbb{R}$ be a real-valued function. The optimization problem can be written as:

$$\min_{x \in X} f(x) \quad \text{subject to} \quad \begin{aligned} g_i(x) &= 0 \quad \forall i \in \{1, \ldots, p\} \\ h_i(x) &\leq 0 \quad \forall i \in \{1, \ldots, m\} \end{aligned}$$

- f is the objective or cost function.
- $g_i : X \to \mathbb{R}$ characterize equality constraints.
- $h_i : X \to \mathbb{R}$ characterize inequality constraints.
- x is **feasible** if $x \in X$ and it satisfies the constraints.

Sufficient conditions for existence and uniqueness

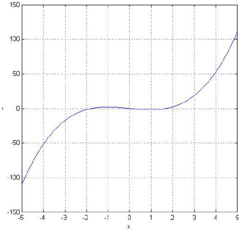
- Extreme Value Theorem: Let $f : X \to \mathbb{R}$ be a upper-semicontinuous function and X be a compact set. Then $\exists x, y \in X$ such that $f(x) = \sup f(S)$.
- Sufficient, but not necessary conditions for existence of optima.
- Let D be a nonempty,convex subset of $\mathbb{R}^n$,and $f : D \to \mathbb{R}$ is strictly quasiconcave. Then if f attains a maximum on D, then the solution is unique.
- Sufficient, but not necessary conditions for the uniqueness of optima.

Locally Optimal Points

- x is locally optimal if there is an $R > 0$ such that x is optimal for

$$\min_{z \in X} f(z) \quad \text{subject to} \quad \begin{aligned} g_i(z) &= 0 \quad \forall i \in \{1, \ldots, p\} \\ h_i(z) &\leq 0 \quad \forall i \in \{1, \ldots, m\} \\ ||z - x|| &\leq R \end{aligned}$$

- $f(x) = x^3 - 3x$: local optimum at $x = 1$



Necessary conditions for (interior) local optima

- Let S be a subset of $\mathbb{R}^n$, and let f be a real-valued function defined on S. If $x^* \in \text{int}S$ is a local maximizer or minimizer of f, and f is differentiable at x^*, then

$$\frac{\partial f(x^*)}{\partial x_j} = 0$$

  for all $j = 1, 2, \ldots n$.
- Then finding the global maximizer boils down to
  1. finding all (interior) local maximizer;
  2. Finding boundary points at which the value of function is greatest;
  3. Among all the points from step 1 and 2, find the points at which f is the greatest.

F.O.C and global optima

Let T be a convex subset of $\mathbb{R}^n$, and let $f : T \to \mathbb{R}$ be differentiable. If $x^* \in \operatorname{int} T$, then

- if f is concave then x^* is a global maximizer of f in T if and only if
$$\frac{\partial f}{\partial x_j}(x^*) = 0$$
for all $j = 1, 2, ...n$ and

- if f is convex then x^* is a global minimizer of f in T if and only if
$$\frac{\partial f}{\partial x_j}(x^*)$$
for all $j = 1, 2, ...n$.

Constrained optimization: the Lagrangian

- Standard-form problem:

$$\max_{x \in X} f(x) \text{ subject to } \begin{aligned} g_i(x) &= 0 \quad \forall i \in \{1, \ldots, p\} \\ h_i(x) &\leq 0 \quad \forall i \in \{1, \ldots, m\} \end{aligned}$$

- **Lagrangian**: $L : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$

$$L(x, \nu, \lambda) = f(x) + \sum_{i=1}^{p} \nu_i g_i(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

- f is a linear combination of the constraints
- λ_i is the Lagrange multiplier associated with $h_i(x) \leq 0$.
- ν_i is the Lagrange multiplier associated with $g_i(x) = 0$.

Implicit Function Theorem

- Let $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ be a continuously differentiable function. Take $(x_0, y_0) \in \mathbb{R}^{n+m}$ and $z_0 \in \mathbb{R}^m$ such that $f(x_0, y_0) = z_0$. If the Jacobian matrix $D_y f$ is invertible at (x_0, y_0), then there is an open set $S \subseteq \mathbb{R}^n$ containing x_0 such that there exists a function $g : S \to \mathbb{R}^m$ where $g(x_0) = y_0$ and $\forall x \in S$, $f(x, g(x)) = z_0$. Moreover, we have:

$$D_x g = -[D_y f]^{-1} D_x f$$

- When $n = m = 1$, the theorem can be simplified to the following:
    - Let $f(x, y) = c$. If $\frac{\partial f}{\partial y} \neq 0$, then the following derivative exists in some neighborhood of x:

$$\frac{\partial y}{\partial x} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}}$$

Necessity, equality constraints

- Let S be an open subset of $\mathbb{R}^n$, and let $f : S \to \mathbb{R}$ and $g_j : S \to \mathbb{R}, j = 1, ..., m$ be continuously differentiable functions with $m \leq n$. If $x^* \in S$ and it is a solution to problem:

$$\max_{x \in X} f(x) \text{ subject to } \quad g_i(x) = 0 \quad \forall i \in \{1, \ldots, m\}$$

and suppose the Jacobian matrix of $(g_1, ...g_m)$ at the point x^* has m linearly independent columns(that is, the rank of the matrix is m). Then there exists a unique $\lambda^* = (\lambda_1^*, ..., \lambda_m^*)$such that x^* satisfies the first order conditions

$$\frac{\partial L}{\partial x_k}(x^*, \lambda^*) = \frac{\partial f}{\partial x_k}(x^*) + \sum_{i=1}^{m} \lambda_j^* \frac{\partial g_j}{\partial x_k}(x^*) = 0$$

for $k = 1, ...n$ and

$$\frac{\partial L}{\partial \lambda_j}(x^*, \lambda^*) = g_j(x^*) = 0$$

Necessity, inequality constraints

- Let S be an open subset of $\mathbb{R}^n$, and let $f : S \to \mathbb{R}$ and $g_j : S \to \mathbb{R}, j = 1, ..., m$ be continuously differentiable functions with $m \leq n$. If $x^* \in S$ and it is a solution to problem:

$$\max_{x \in X} f(x) \ \text{ subject to } \quad g_i(x) \geq 0 \quad \forall i \in \{1, \dots, m\}$$

and suppose at least one of the CQs is satisfied. Then there exists $\lambda^* = (\lambda_1^*, ..., \lambda_m^*) \in \mathbb{R}_+^m$such that x^* satisfies the first order conditions

$$\frac{\partial L}{\partial x_k}(x^*, \lambda^*) = \frac{\partial f}{\partial x_k}(x^*) + \sum_{i=1}^{m} \lambda_j^* \frac{\partial g_j}{\partial x_k}(x^*) = 0$$

for $k = 1, ...n$ and

$$\lambda_j^* g_j(x^*) = 0, \ g_j(x^*) \geq 0$$

for $j = 1, ...m$.

Constraint qualifications

- Let $(g_1^B, ...g_q^B)$ be the constraints that are binding at x^*, with $q \leq m$. then $a \leq n$, and the Jacobian matrix of $(g_1^B, ...g_q^B)$ at the point x^* has q linearly independent columns.
- $g_i(x)$ is convex for $j = 1, ..., m$.
- $g_i(x)$ is concave for $i = 1, ..., m$ and Slater's condition holds:

$$\exists x \in int(X) : h_i(x) < 0, \; i = 1, \ldots, m,$$

Sufficiency, inequality constraints

- Let $S \subset \mathbb{R}^n$ and let $f : S \to \mathbb{R}$ and $g_j : S \to \mathbb{R}$, $j = 1, \ldots m$ be continuously differentiable function. Suppose that f is concave and
  - Either g_j is affine for each $j = 1, \ldots m$
  - or g_j is concave for each $j = 1, \ldots m$ and s Slater's condition holds.

  then x^* solves the problem if and only if there exists $\lambda \in \mathbb{R}^m_+$, such that

$$\frac{\partial L}{\partial x_k}(x^*, \lambda^*) = \frac{\partial f}{\partial x_k}(x^*) + \sum_{i=1}^{m} \lambda_j^* \frac{\partial g_j}{\partial x_k}(x^*) = 0$$

  for $k = 1, \ldots n$ and

$$\lambda_j^* g_j(x^*) = 0, \ g_j(x^*) \geq 0$$

  for $j = 1, \ldots m$.

Karush-Kuhn-Tucker (KKT) Conditions

Let S be an open subset of $\mathbb{R}^n$, and let $f : S \to \mathbb{R}$ and $g_j : S \to \mathbb{R}, j = 1, ..., m$ be continuously differentiable functions with $m \leq n$.

- When at least one of the constraint qualifications is satisfied, if $x^* \in X$ is optimal, then there must exist $\nu \in \mathbb{R}^p$ and $\lambda \in \mathbb{R}^m$ such that:
  - $\nabla f(x^*) + \sum_{i=1}^{p} \nu_i \nabla g_i(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0$
  - $g_i(x^*) = 0 \quad \forall i \in \{1, ..., p\}$
  - $h_i(x^*) \leq 0 \quad \forall i \in \{1, ..., m\}$
  - $\lambda_i \geq 0 \quad \forall i \in \{1, ..., m\}$
  - $\lambda_i h_i(x^*) = 0 \quad \forall i \in \{1, ..., m\}$
- KKT conditions are necessary, but not generally sufficient.
- KKT conditions are sufficient for convex problems.

Envelope Theorem

- Consider our problem with an additional parameter t:

$$\min_{x \in X} f(x,t) \text{ subject to } \begin{aligned} g_i(x,t) &= 0 \quad \forall i \in \{1,\ldots,p\} \\ h_i(x,t) &\leq 0 \quad \forall i \in \{1,\ldots,m\} \end{aligned}$$

- Let $V(t)$ denote the minimized value of the above problem.
- The associated Lagrangian is:

$$\mathcal{L}(x,\lambda,\nu,t) = f(x,t) + \sum_{i=1}^{m} \nu_i g_i(x,t) + \sum_{j=1}^{n} \lambda_i h_i(x,t)$$

- Envelope theorem states that:

$$\frac{dV(t_1)}{dt} = \frac{\partial \mathcal{L}(x^*(t_1), \lambda^*(t_1), \nu^*(t_1), t = t_1)}{\partial t}$$

Calculus of Variations

- A general formulation for these problems is as follows:

$$\max_{x(t)} \int_{t_0}^{t_1} f(t, x(t), x'(t))dt \quad \text{subject to} \quad x(t_0) = x_0, x(t_1) = x_1$$

- Here $f : [t_0, t_1] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is smooth with respect to the second and third arguments and continuous in the first argument.

- Assume there exists a function $x^*(t)$ solving the above problem. Then for any $\varepsilon > 0$ and any continuously differentiable function $h : [t_0, t_1] \to \mathbb{R}$ with $h(t_0) = h(t_1) = 0$:

$$\int_{t_0}^{t_1} f(t, x^*(t), x^{*'}(t))dt \geq \int_{t_0}^{t_1} f(t, x^*(t) + \varepsilon h(t), x^{*'}(t) + \varepsilon h'(t))dt$$

- The derivative with respect to ε must be zero when $\varepsilon = 0$:

$$\int_{t_0}^{t_1} [f_x(t, x^*(t), x^{*'}(t))h(t) + f_{x'}(t, x^*(t), x^{*'}(t))h'(t)]dt = 0$$

Calculus of Variations

- Applying method of integration by parts to $\int f_{x'} h' dt$ where $u = f_{x'}$ and $dv = h' dt$, we get:

$$\int_{t_0}^{t_1} f_{x'} h' dt = f_{x'} h \Big|_{t_0}^{t_1} - \int_{t_0}^{t_1} h \frac{df_{x'}}{dt} dt$$

- Since $h(t_0) = h(t_1) = 0$, plugging this back we get:

$$\int_{t_0}^{t_1} \left[f_x(t, x^*(t), x^{*'}(t)) - \frac{df_{x'}(t, x^*(t), x^{*'}(t))}{dt} \right] h(t) dt = 0$$

- But since this is true for every $h(t)$, we must have:

$$f_x(t, x^*(t), x^{*'}(t)) = \frac{df_{x'}(t, x^*(t), x^{*'}(t))}{dt} \quad \text{(Euler equation)}$$

Calculus of Variations: Example

- Find the shortest path $y(x)$ between two points (x_1, y_1) and (x_2, y_2).
- Recall that $\int \sqrt{1+(y'(x))^2}dx$ is the length of the path between the points. Consider $f(y'(x)) = \sqrt{1+(y'(x))^2}$. So we have the following formulation

$$\min \int_{x_1}^{x_2} f(y'(x))dx \quad \text{or} \quad \min \int_{x_1}^{x_2} \sqrt{1+(y'(x))^2}dx \quad .$$

- Applying the Euler equation we get: (Note that $f_y = 0$.)

$$\frac{y'(x)}{\sqrt{1+(y'(x))^2}} = c \Rightarrow y(x) = \frac{c}{\sqrt{1-c^2}}x + d$$

Appendix A: Differentiation

- Let $f : (a, b) \to \mathbb{R}$. Define the quotient function as

$$\phi(t) = \frac{f(t) - f(x)}{t - x}$$

- $f'(x^-)$ is called the left derivative at x if $f'(x^-) := \lim_{t \to x-} \phi(t)$ exists;
- $f'(x^+)$ is called the right derivative at x if $f'(x^+) := \lim_{t \to x-} \phi(t)$ exists;
- f is differentiable at x if $\lim_{t \to x} \phi(t)$ exists, which is denoted as $f'(x)$, it is differentiable at on (a, b) if it is differentiable everywhere on (a, b)
- say f is continuously differentiable (or $f \in C^1$) on (a, b) if f' is continuous.
- Differentiablity implies continuity, but the inverse is not true.

Appendix B: Jacobian Matrix

- Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable function. The **Jacobian matrix** of f is an $m \times n$ matrix of first derivatives:

$$D_x f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Appendix C: Riemann integration

- Riemann Integral is the limit of partial sums:

$$\int_a^b f(x)dx = \lim_{h\to 0} \sum_{i=1}^{\frac{b-a}{h}} f(a+ih)h$$

- All $C([a,b])$ are integrable;
- **The fundamental theorem of calculus**: For any $f \in C[a,b]$ and $f : [a,b] \to \mathbb{R}$, we have

$$F(x) = F(a) + \int_a^x f(x)dx, \quad \forall x \in [a,b]$$

if and only if $F \in C^1[a,b]$ and $F' = f$.
- One important method to compute integrals is **integration by parts**:

$$\int u\frac{dv}{dx}dx = uv - \int v\frac{du}{dx}dx$$