5. Differentiation II

Daisuke Oyama

Mathematics II

April 26, 2024

Vectors and Matrices

- We regard elements in $\mathbb{R}^N$ as column vectors.
- We denote the set of $M \times N$ matrices by $\mathbb{R}^{M\times N}$,

$$\begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{MN} \end{pmatrix} \in \mathbb{R}^{M\times N}.$$

- For $A \in \mathbb{R}^{M\times N}$, $A^{\mathrm{T}} \in \mathbb{R}^{N\times M}$ denotes the transpose of A.
- $\mathbb{R}^N$ and $\mathbb{R}^{N\times 1}$ are naturally identified,
  and we use the natural identification

$$\underbrace{x \cdot y}_{\text{real number}} = \underbrace{x^{\mathrm{T}} y}_{1 \times 1 \text{ matrix}}$$

for $x, y \in \mathbb{R}^N$ or $x, y \in \mathbb{R}^{N\times 1}$.

Little o Notation

- For functions $f, g\colon U \to \mathbb{R}$,
  where $U \subset \mathbb{R}^N$ is an open neighborhood of $\bar{x} \in \mathbb{R}^N$,
  if $\lim_{x \to \bar{x}} g(x) = 0$ and $\lim_{x \to \bar{x}} \frac{f(x)}{g(x)} = 0$, we write

  $$f(x) = o(g(x)) \text{ as } x \to \bar{x}.$$

- By $f(x) = h(x) + o(g(x))$, we mean $f(x) - h(x) = o(g(x))$.

Partial Differentiation

Let U be a nonempty open subset of $\mathbb{R}^N$.

- A function $f: U \to \mathbb{R}$ is *partially differentiable with respect to* x_i at $\bar{x} \in U$ if the function $x_i \mapsto f(\bar{x}_1, \ldots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \ldots, \bar{x}_N)$ is differentiable at $\bar{x}_i$.
- In this case, the differential coefficient is denoted by $f_i(\bar{x})$, $f_{x_i}(\bar{x})$, or $\frac{\partial f}{\partial x_i}(\bar{x})$,
  and is called the *partial differential coefficient* of f with respect to x_i at $\bar{x}$.

  We also say that $\frac{\partial f}{\partial x_i}(\bar{x})$ exists.
- f is partially differentiable with respect to x_i if it is partially differentiable with respect to x_i at all $\bar{x} \in U$.
- The function $x \mapsto f_i(x)$ is called the *partial derivative function* (or *partial derivative*) of f with respect to x_i and is denoted by f_i, f_{x_i}, or $\frac{\partial f}{\partial x_i}$.

Gradient Vectors

Let U be a nonempty open subset of $\mathbb{R}^N$.

- For a function $f\colon U \to \mathbb{R}$, if $\frac{\partial f}{\partial x_i}(\bar{x})$ exists for all $i = 1, \ldots, N$, we write

$$\nabla f(\bar{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\bar{x}) \\ \vdots \\ \frac{\partial f}{\partial x_N}(\bar{x}) \end{pmatrix} \in \mathbb{R}^N,$$

which is called the *gradient vector* (or *gradient*) of f at $\bar{x}$.

Jacobian Matrices

Let U be a nonempty open subset of $\mathbb{R}^N$.

- For a function $f\colon U \to \mathbb{R}^M$, if $\frac{\partial f_j}{\partial x_i}(\bar{x})$ exists for all $i = 1, \dots, N$ and $j = 1, \dots, M$, we write

$$Df(\bar{x}) = \begin{pmatrix} \nabla f_1(\bar{x})^{\mathrm{T}} \\ \vdots \\ \nabla f_M(\bar{x})^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\bar{x}) & \cdots & \frac{\partial f_1}{\partial x_N}(\bar{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial x_1}(\bar{x}) & \cdots & \frac{\partial f_M}{\partial x_N}(\bar{x}) \end{pmatrix} \in \mathbb{R}^{M \times N},$$

  which is called the *Jacobian matrix* (or *Jacobian*) of f at $\bar{x}$.
- For a function $f\colon U \to \mathbb{R}$, $Df(\bar{x}) = \nabla f(\bar{x})^{\mathrm{T}}$.

- For a function $f(x, y)$ of $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^S$, we often write

$$D_x f(\bar{x}, \bar{y}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\bar{x}, \bar{y}) & \cdots & \frac{\partial f_1}{\partial x_N}(\bar{x}, \bar{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial x_1}(\bar{x}, \bar{y}) & \cdots & \frac{\partial f_M}{\partial x_N}(\bar{x}, \bar{y}) \end{pmatrix} \in \mathbb{R}^{M \times N},$$

and

$$D_y f(\bar{x}, \bar{y}) = \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(\bar{x}, \bar{y}) & \cdots & \frac{\partial f_1}{\partial y_S}(\bar{x}, \bar{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial y_1}(\bar{x}, \bar{y}) & \cdots & \frac{\partial f_M}{\partial y_S}(\bar{x}, \bar{y}) \end{pmatrix} \in \mathbb{R}^{M \times S},$$

where

$$Df(\bar{x}, \bar{y}) = \begin{pmatrix} D_x f(\bar{x}, \bar{y}) & D_y f(\bar{x}, \bar{y}) \end{pmatrix} \in \mathbb{R}^{M \times (N+S)}.$$

Differentiation in Several Variables

Let U be a nonempty open subset of $\mathbb{R}^N$.

Definition 5.2

A function $f\colon U \to \mathbb{R}$ is *differentiable* (or *totally differentiable*) at $\bar{x} \in U$ if there exists $\bar{p} \in \mathbb{R}^N$ such that

$$\lim_{z \to 0} \frac{f(\bar{x}+z) - f(\bar{x}) - \bar{p} \cdot z}{\|z\|} = 0,$$

or $f(\bar{x}+z) = f(\bar{x}) + \bar{p} \cdot z + o(\|z\|)$ as $z \to 0$,
i.e., for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$0 < \|z\| < \delta,\ \bar{x}+z \in U \Longrightarrow \frac{|f(\bar{x}+z) - f(\bar{x}) - \bar{p} \cdot z|}{\|z\|} < \varepsilon.$$

- In this case, $\frac{\partial f}{\partial x_i}(\bar{x})$ exists for all $i = 1, \ldots, N$, and $\bar{p} = \nabla f(\bar{x})$.

Differentiability, Continuity, Partial Differentiability

Proposition 5.8

If f is differentiable at $\bar{x}$, then it is continuous at $\bar{x}$, and partially differentiable with respect to x_i at $\bar{x}$ for each i.

However,

- partial differentiability does not imply differentiability; and
- partial differentiability does not even imply continuity.

Continuous Differentiability and Differentiability

Let U be a nonempty open subset of $\mathbb{R}^N$.

- $f : U \to \mathbb{R}$ is *continuously differentiable* or *of class C^1* if it is partially differentiable with respect to $x_1, \ldots, x_N$ and its partial derivative functions $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_N}$ are continuous.

Proposition 5.9

If f is continuously differentiable, then it is differentiable.

Vector-Valued Functions

Let U be a nonempty open subset of $\mathbb{R}^N$.

- For a function $f\colon U \to \mathbb{R}^M$, we write $f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_M(x) \end{pmatrix}$.

  f is differentiable if f_m is differentiable for all $m = 1, \dots, M$.

- When f is differentiable,

$$\lim_{z \to 0} \frac{1}{\|z\|}(f(\bar{x} + z) - f(\bar{x}) - Df(\bar{x})z) = 0,$$

  where $Df(\bar{x}) \in \mathbb{R}^{M \times N}$ is the Jacobian matrix of f at $\bar{x}$.

- f is of class C^1 if f_m is of class C^1 for all $m = 1, \dots, M$.

Product Rule

Let $U \subset \mathbb{R}^N$ be a nonempty open set.

Proposition 5.10

Suppose that $f\colon U \to \mathbb{R}^M$ and $g\colon U \to \mathbb{R}^M$ are differentiable.
Define the function $h\colon U \to \mathbb{R}$ by $h(x) = f(x)^{\mathrm{T}} g(x)$.
Then h is differentiable and satisfies

$$\underbrace{Dh(x)}_{1\times N} = \underbrace{g(x)^{\mathrm{T}}}_{1\times M} \underbrace{Df(x)}_{M\times N} + \underbrace{f(x)^{\mathrm{T}}}_{1\times M} \underbrace{Dg(x)}_{M\times N}$$

for all $x \in U$.

Chain Rule

Let $U \subset \mathbb{R}^N$ and $V \subset \mathbb{R}^S$ be nonempty open sets.

Proposition 5.11

Suppose that $g\colon V \to U$ and $f\colon U \to \mathbb{R}^M$ are differentiable.
Define the function $h\colon V \to \mathbb{R}^M$ by $h(x) = f(g(x))$.
Then h is differentiable and satisfies

$$\underbrace{Dh(x)}_{M\times S} = \underbrace{Df(g(x))}_{M\times N}\underbrace{Dg(x)}_{N\times S}$$

for all $x \in V$.

Example 1-1

- For a function $f\colon \mathbb{R}^N \to \mathbb{R}$ and $y, z \in \mathbb{R}^N$, consider the function $h\colon \mathbb{R} \to \mathbb{R}$ defined by $h(\alpha) = f(y + \alpha z)$.
- Define the function $g\colon \mathbb{R} \to \mathbb{R}^N$ by $g(\alpha) = y + \alpha z$.

  Then $h(\alpha) = f(g(\alpha))$.
- By the Chain rule,

$$
\begin{aligned}
h'(\alpha) = Dh(\alpha) &= Df(g(\alpha))Dg(\alpha) \\
&= \underbrace{Df(y+\alpha z)}_{1\times N}\underbrace{z}_{N\times 1} \qquad \text{(matrix product)} \\
&= \underbrace{\nabla f(y+\alpha z)}_{\in\mathbb{R}^N}\cdot\underbrace{z}_{\in\mathbb{R}^N}. \qquad \text{(inner product)}
\end{aligned}
$$

Example 1-2

- For a function $f\colon \mathbb{R}^N \to \mathbb{R}^N$ and $y, z \in \mathbb{R}^N$, consider the function $k\colon \mathbb{R} \to \mathbb{R}$ defined by

$$k(\alpha) = z^{\mathrm{T}} f(y + \alpha z).$$

- By the Chain rule,

$$\begin{aligned} k'(\alpha) = Dk(\alpha) &= z^{\mathrm{T}} D_\alpha\big[f(y + \alpha z)\big] \\ &= \underbrace{z^{\mathrm{T}}}_{1\times N} \underbrace{Df(y + \alpha z)}_{N\times N} \underbrace{z}_{N\times 1}. \end{aligned}$$

Example 2: Slutsky Equation

- - $x\colon \mathbb{R}^N_{++} \times \mathbb{R}_+ \to \mathbb{R}^N_+$: Walrasian demand function
  - $h\colon \mathbb{R}^N_{++} \times \mathbb{R} \to \mathbb{R}^N_+$: Hicksian demand function
  - $e\colon \mathbb{R}^N_{++} \times \mathbb{R} \to \mathbb{R}$: expenditure function
- By duality, we have $h(p) = x(p, e(p))$.

  (The fixed utility level u is omitted.)

  I.e., if $g\colon \mathbb{R}^N_{++} \to \mathbb{R}^N_{++} \times \mathbb{R}_+$ is defined by $g(q) = (q, e(q))$, then $h(p) = x(g(p))$.
- $Dg(q) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ e_1 & e_2 \end{pmatrix}$, where $e_n = \dfrac{\partial e}{\partial p_n}$ (and $N = 2$).
- We will also write $x_{np_k} = \dfrac{\partial x_n}{\partial p_k}$ and $x_{nw} = \dfrac{\partial x_n}{\partial w}$.

Then by the Chain Rule,

$$
\begin{aligned}
Dh(p) &= Dx(g(p))Dg(p) \\
&= \begin{pmatrix} x_{1p_1} & x_{1p_2} & x_{1w} \\ x_{2p_1} & x_{2p_2} & x_{2w} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ e_1 & e_2 \end{pmatrix} \\
&= \begin{pmatrix} x_{1p_1} & x_{1p_2} \\ x_{2p_1} & x_{2p_2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} x_{1w} \\ x_{2w} \end{pmatrix} \begin{pmatrix} e_1 & e_2 \end{pmatrix} \\
&= \underbrace{D_p x(p, e(p))}_{N \times N} + \underbrace{D_w x(p, e(p))}_{N \times 1} \underbrace{D_p e(p)}_{1 \times N} \\
&= \underbrace{D_p x(p, e(p))}_{N \times N} + \underbrace{D_w x(p, e(p))}_{N \times 1} \underbrace{h(p)^{\mathrm{T}}}_{1 \times N},
\end{aligned}
$$

where the last equality follows from $\underbrace{\nabla e(p)}_{N \times 1} = \underbrace{h(p)}_{N \times 1}$

("Hotelling's Lemma").

Example 3: Homogeneous Functions and Euler's Formula

Definition 5.3

A function $f\colon \mathbb{R}^N_+ \to \mathbb{R}$ is *homogeneous of degree* k if

$$f(tx) = t^k f(x)$$

for all $t > 0$ and all $x \in \mathbb{R}^N_+$.

Proposition 5.12

If f is homogeneous of degree k and differentiable, then for all i, $\frac{\partial f}{\partial x_i}$ is homogeneous of degree $k-1$.

Proof

- Since $f(tx) = t^k f(x)$ holds for any value of x_i, it holds that $\frac{\partial}{\partial x_i}$(LHS) $= \frac{\partial}{\partial x_i}$(RHS).
- Since

$$\frac{\partial}{\partial x_i}(\mathsf{LHS}) = t\frac{\partial f}{\partial x_i}(tx),$$

and

$$\frac{\partial}{\partial x_i}(\mathsf{RHS}) = t^k\frac{\partial f}{\partial x_i}(x),$$

we have $\frac{\partial f}{\partial x_i}(tx) = t^{k-1}\frac{\partial f}{\partial x_i}(x)$.

Proposition 5.13

If f is homogeneous of degree k and differentiable, then

$$\nabla f(x) \cdot x = kf(x)$$

for all $x \in \mathbb{R}^N_+$.

Proof

- Since $f(tx) = t^k f(x)$ holds for any value of t, it holds that $\frac{\partial}{\partial t}(\text{LHS}) = \frac{\partial}{\partial t}(\text{RHS})$.
- We have

  $$\frac{\partial}{\partial t}(\text{LHS}) = \nabla f(tx) \cdot x,$$

  and

  $$\frac{\partial}{\partial t}(\text{RHS}) = kt^{k-1} f(x).$$

  Since these are equal, evaluating at $t = 1$ we have $\nabla f(x) \cdot x = kf(x)$.

Example 4: A Property of the Hicksian Demand Function

- The Hicksian demand function $h(p, u)$ is homogeneous of degree 0 in p.
- By Proposition 5.13, we have

$$\underbrace{D_p h(p, u)}_{N \times N} \underbrace{p}_{N \times 1} = \underbrace{0}_{N \times 1}.$$

Mean Value Theorem in Several Variables

Let $U \subset \mathbb{R}^N$ be a nonempty open convex set.

Proposition 5.14

Suppose that $f : U \to \mathbb{R}$ is differentiable.
Then for any $x, y \in U$, there exists $\alpha_0 \in (0, 1)$ such that

$$f(y) - f(x) = \nabla f((1 - \alpha_0)x + \alpha_0 y) \cdot (y - x).$$

Proof

- Consider the differentiable function $h(\alpha) = f(x + \alpha(y - x))$.
- By the Mean Value Theorem in one variable, there exists $\alpha_0 \in (0, 1)$ such that $h(1) - h(0) = h'(\alpha_0)(1 - 0)$, or $f(y) - f(x) = \nabla f(x + \alpha_0(y - x)) \cdot (y - x)$.

Second Order Differentiation

- The partial derivative of $\frac{\partial f}{\partial x_i}$ with respect to x_i is written as

$$\frac{\partial^2 f}{\partial x_i^2} \quad \text{or} \quad f_{x_i x_i} \quad \text{or} \quad f_{ii}.$$

- The partial derivative of $\frac{\partial f}{\partial x_i}$ with respect to x_j is written as

$$\frac{\partial^2 f}{\partial x_j \partial x_i} \quad \text{or} \quad f_{x_i x_j} \quad \text{or} \quad f_{ij}.$$

- These are called the second partial derivative functions, or second partial derivatives, of f.
- f is *twice continuously differentiable* or *of class* C^2 if all the second partial derivatives exist and are continuous.

Hessian Matrices

Let U be a nonempty open subset of $\mathbb{R}^N$.

- For a function $f\colon U \to \mathbb{R}$, if all the second partial derivatives exist at $\bar{x}$, we write

$$
\begin{aligned}
D^2 f(\bar{x}) &= D\nabla f(\bar{x}) \\
&= \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(\bar{x}) & \cdots & \dfrac{\partial^2 f}{\partial x_N \partial x_1}(\bar{x}) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_N}(\bar{x}) & \cdots & \dfrac{\partial^2 f}{\partial x_N^2}(\bar{x}) \end{pmatrix} \in \mathbb{R}^{N\times N},
\end{aligned}
$$

  which is called the *Hessian matrix* (or *Hessian*) of f at $\bar{x}$.
- Some textbooks define the Hessian to be the transpose of this matrix.

Young's Theorem

- In general, $\frac{\partial^2 f}{\partial x_j \partial x_i}(x) \neq \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$.

Proposition 5.15

If $f\colon U \to \mathbb{R}$ is of class C^2, then $D^2 f(x)$ is symmetric, i.e.,

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \text{ for all } i, j = 1, \dots, N,$$

for all $x \in U$.

- There are other, weaker conditions,
  such as "all the first partial derivatives are differentiable".
- The above proposition, or one with a weaker condition, is called Young's theorem, Schwarz's theorem, or Clairaut's theorem.

Example 5: Symmetry of $D_p h(p, u)$

- By "Hotelling's Lemma", $h(p, u) = \nabla_p e(p, u)$.
- If h is of class C^1 in p, so that e is of class C^2 in p, then $D_p h(p, u) = D^2 e(p, u)$ is symmetric by Young's Theorem.

Example 6

- For a function $f\colon \mathbb{R}^N \to \mathbb{R}$ and $y, z \in \mathbb{R}^N$,
  define the function $g\colon \mathbb{R} \to \mathbb{R}^N$ by $g(\alpha) = \nabla f(y + \alpha z)$.

  Then by the Chain rule,

  $$Dg(\alpha) = D\nabla f(y + \alpha z)z = \underbrace{D^2 f(y + \alpha z)}_{N\times N}\underbrace{z}_{N\times 1} \in \mathbb{R}^{N\times 1}.$$

- Consider the function $h\colon \mathbb{R} \to \mathbb{R}$ defined by $h(\alpha) = f(y+\alpha z)$.

  As we have seen $h'(\alpha) = \nabla f(y + \alpha z) \cdot z = g(\alpha) \cdot z$.

  Then,

  $$\begin{aligned} h''(\alpha) &= Dg(\alpha) \cdot z \\ &= \big(D^2 f(y + \alpha z)z\big) \cdot z = z \cdot D^2 f(y + \alpha z)z \\ &= z^{\mathrm{T}} D^2 f(y + \alpha z)z. \end{aligned}$$

Taylor's Theorem: 2nd Order Case

Let $U \subset \mathbb{R}^N$ be a nonempty open convex set.
Let $\bar{x} \in U$ and let $z \in \mathbb{R}^N$ such that $x + z \in U$.

Proposition 5.16

1. *If* $f\colon U \to \mathbb{R}$ *is differentiable and* ∇f *is differentiable at* $\bar{x} \in U$, *then*

$$f(\bar{x} + z) = f(\bar{x}) + \nabla f(\bar{x}) \cdot z + \frac{1}{2} z \cdot D^2 f(\bar{x}) z + o(\|z\|^2).$$

2. *If* f *is twice differentiable, then there exists* $\alpha_0 \in (0, 1)$ *such that*

$$f(\bar{x} + z) = f(\bar{x}) + \nabla f(\bar{x}) \cdot z + \frac{1}{2} z \cdot D^2 f(\bar{x} + \alpha_0 z) z.$$

Implicit Function Theorem

Let $A \subset \mathbb{R}^N$ and $B \subset \mathbb{R}^M$ be nonempty open sets.

Proposition 5.17

Suppose that $f\colon A \times B \to \mathbb{R}^N$, $(x, q) \mapsto f(x, q)$, is of class C^1. Assume that $f(\bar{x}, \bar{q}) = 0$, where $(\bar{x}, \bar{q}) \in A \times B$, and $|D_x f(\bar{x}, \bar{q})| \neq 0$.
Then there exist an open neighborhood $U \subset A$ of $\bar{x}$, an open neighborhood $V \subset B$ of $\bar{q}$, and a C^1 function $\eta\colon V \to U$ that satisfy the following:

- *for all $(x, q) \in U \times V$, $f(x, q) = 0 \iff x = \eta(q)$; and*
- $D\eta(\bar{q}) = -[D_x f(\bar{x}, \bar{q})]^{-1} D_q f(\bar{x}, \bar{q})$.

Intuition

- Suppose that $f(\bar{x}, \bar{q}) = 0$.
- Given $q \approx \bar{q}$, we want to solve the equation $f(x, q) = 0$ in x.
- Locally, the equation is approximated by the *linear* equation

$$\underbrace{D_x f(\bar{x}, \bar{q})}_{N \times N} \underbrace{(x - \bar{x})}_{\in \mathbb{R}^N} + \underbrace{D_q f(\bar{x}, \bar{q})}_{N \times M} \underbrace{(q - \bar{q})}_{\in \mathbb{R}^M} = \underbrace{0}_{\in \mathbb{R}^N}.$$

- If $|D_x f(\bar{x}, \bar{q})| \neq 0$, then this linear equation has a solution, and the solution is given as a function of q by

$$\theta(q) = \bar{x} - [D_x f(\bar{x}, \bar{q})]^{-1} D_q f(\bar{x}, \bar{q})(q - \bar{q}),$$

  where

$$D\theta(q) = -[D_x f(\bar{x}, \bar{q})]^{-1} D_q f(\bar{x}, \bar{q}).$$

- $\theta(q)$ is a linear approximation of the solution $\eta(q)$ of the original equation.

Concave Functions

Definition 5.4

Let $X \subset \mathbb{R}^N$ be a nonempty convex set.

- A function $f\colon X \to \mathbb{R}$ is *concave* if

  $$f((1-\alpha)x + \alpha x') \geq (1-\alpha)f(x) + \alpha f(x')$$

  for all $x, x' \in X$ and all $\alpha \in [0,1]$.

- $f\colon X \to \mathbb{R}$ is *strictly concave* if

  $$f((1-\alpha)x + \alpha x') > (1-\alpha)f(x) + \alpha f(x')$$

  for all $x, x' \in X$ with $x \neq x'$ and all $\alpha \in (0,1)$.

- $f\colon X \to \mathbb{R}$ is *convex* (*strictly convex*, resp.) if $-f$ is concave (strictly concave, resp.).

Characterization of Concave Functions

Let $X \subset \mathbb{R}^N$ be a nonempty convex set.

Lemma 5.18

$f\colon X \to \mathbb{R}$ is (strictly) concave if and only if for any $x \in X$ and any $z \in \mathbb{R}^N$ with $x + z \in X$, for $t \in (0, 1]$,

$$\frac{f(x + tz) - f(x)}{t}$$

is nonincreasing (strictly decreasing) in t.

Characterization via Gradient

Let $X \subset \mathbb{R}^N$ be a nonempty open convex set.

Proposition 5.19

Suppose that $f : X \to \mathbb{R}$ is differentiable.

- *f is concave if and only if*

  $$f(x+z) \leq f(x) + \nabla f(x) \cdot z$$

  for all $x \in X$ and all $z \in \mathbb{R}^N$ with $x + z \in X$.

- *f is strictly concave if and only if*

  $$f(x+z) < f(x) + \nabla f(x) \cdot z$$

  for all $x \in X$ and all $z \neq 0$ with $x + z \in X$.

Proof (1/2)

- The "if" part:

  Take any $x, x' \in X$ and $\alpha \in (0,1)$, and denote $x'' = (1-\alpha)x + \alpha x'$. By assumption,

  $$f(x) \leq f(x'') + \nabla f(x'') \cdot (x - x''), \tag{1}$$
  $$f(x') \leq f(x'') + \nabla f(x'') \cdot (x' - x''). \tag{2}$$

  From (1) $\times (1-\alpha) +$ (2) $\times \alpha$, we have

  $$(1-\alpha) f(x) + \alpha f(x') \leq f(x'').$$

- For strict concavity, replace "$\leq$" with "$<$" (assuming $x \neq x'$).

Proof (2/2)

- The "only if" part: Suppose that f is concave, and fix any $x \in X$ and $z \in \mathbb{R}^N$ with $x + z \in X$.
- By Lemma 5.18, for $t > 0$, $\dfrac{f(x+tz) - f(x)}{t}$ is decreasing in t.
- In particular, we have $\dfrac{f(x+tz) - f(x)}{t} \geq f(x+z) - f(x)$ for $t \in (0,1]$.
- Let $t \searrow 0$. Then by the definition of differentiation,

$$\text{(LHS)} \nearrow \frac{\partial}{\partial t} f(x+tz)\Big|_{t=0} = \nabla f(x+tz) \cdot z\Big|_{t=0} = \nabla f(x) \cdot z.$$

- For strict concavity, replace "$\geq$" with "$>$" (assuming $z \neq 0$).

Characterization via Gradient

Let $X \subset \mathbb{R}^N$ be a nonempty open convex set.

Proposition 5.20

Suppose that $f\colon X \to \mathbb{R}$ is differentiable.

- *f is concave if and only if*

  $$(\nabla f(x') - \nabla f(x)) \cdot (x' - x) \leq 0$$

  for all $x, x' \in X$.

- *f is strictly concave if and only if*

  $$(\nabla f(x') - \nabla f(x)) \cdot (x' - x) < 0$$

  for all $x, x' \in X$ with $x \neq x'$.

Proof (1/2)

- The "if" part:

  Fix any $x \in X$ and $z \in \mathbb{R}^N$ with $x + z \in X$.

- Let

$$g(t) = f(x + tz) - f(x) - \nabla f(x) \cdot (tz).$$

  By Proposition 5.19, it suffices to show that $g(1) \leq 0$.

- For all $t \in (0, 1]$, we have

$$\begin{aligned} g'(t) &= \nabla f(x + tz) \cdot z - \nabla f(x) \cdot z \\ &= (\nabla f(x + tz) - \nabla f(x)) \cdot (tz)/t \leq 0 \end{aligned}$$

  by assumption.

- Since $g(0) = 0$, it follows that $g(1) \leq 0$.
- For strict concavity, replace "$\leq$" with "$<$" (assuming $z \neq 0$).

Proof (2/2)

- The "only if" part:

  Suppose that f is concave, and fix any $x, x' \in X$.

- By Proposition 5.19, we have

$$f(x') \leq f(x) + \nabla f(x) \cdot (x' - x),$$
$$f(x) \leq f(x') + \nabla f(x') \cdot (x - x').$$

- Combining these inequalities, we have

$$0 \leq -(\nabla f(x) - \nabla f(x')) \cdot (x' - x).$$

- For strict concavity, replace "$\leq$" with "$<$" (assuming $x \neq x'$).

Differentiability and Partial Differentiability

Let $X \subset \mathbb{R}^N$ be a nonempty convex set.

Fact 1

Suppose that $f\colon X \to \mathbb{R}$ is concave, and let $\bar{x} \in \operatorname{Int} X$.
If $\frac{\partial f}{\partial x_i}(\bar{x})$ exists for all $i = 1, \dots, N$, then f is differentiable at $\bar{x}$.

- This does not hold for general functions.

Characterization via Hessian

Let $X \subset \mathbb{R}^N$ be a nonempty open convex set.

Proposition 5.21

Suppose that $f\colon X \to \mathbb{R}$ is differentiable and ∇f is differentiable.

- *f is concave if and only if*
  for all $x \in X$, $D^2 f(x)$ is negative semi-definite, i.e.,

  $$z \cdot D^2 f(x) z \leq 0$$

  for all $z \in \mathbb{R}^N$.

- *If for all $x \in X$, $D^2 f(x)$ is negative definite, i.e.,*

  $$z \cdot D^2 f(x) z < 0$$

  for all $z \neq 0$, then f is strictly concave.

Proof (1/2)

- The “if” part: Fix any $x, x' \in X$, and write $z = x' - x$.
- Let

$$g(t) = (\nabla f(x + tz) - \nabla f(x)) \cdot z.$$

  By Proposition 5.20, it suffices to show that $g(1) \leq 0$.
- For all $t \in (0, 1]$, we have

$$g'(t) = z \cdot D^2 f(x + tz) z \leq 0$$

  by assumption.
- Since $g(0) = 0$, it follows that $g(1) \leq 0$.
- For strict concavity, replace “$\leq$” with “$<$” (assuming $x \neq x'$).

Proof (2/2)

- The "only if" part:
  Suppose that f is concave.

  By Proposition 5.20,
  $(\nabla f(x') - \nabla f(x)) \cdot (x' - x) \leq 0$ for any $x, x' \in X$.

- Fix any $x \in X$ and $z \in \mathbb{R}^N$, and consider the function

  $$g(t) = \nabla f(x + tz) \cdot z$$

  (defined for t such that $x + tz \in X$).

- By assumption, for $t' > t$, we have

  $$\begin{aligned}&(g(t') - g(t))(t' - t)\\ &= (\nabla f(x+t'z) - \nabla f(x+tz)) \cdot \{(x+t'z) - (x+tz)\} \leq 0,\end{aligned}$$

  which implies that g is nonincreasing.

- Therefore, $g'(t) = z \cdot D^2 f(x + tz)z \leq 0$ for all t.

  In particular, we have $g'(0) = z \cdot D^2 f(x)z \leq 0$.

Quasi-Concave Functions

Definition 5.5

Let $X \subset \mathbb{R}^N$ be a nonempty convex set.

- $f : X \to \mathbb{R}$ is *quasi-concave* if
  $f((1-\alpha)x + \alpha x') \geq f(x)$
  for all $x, x' \in A$ such that $f(x') \geq f(x)$ and all $\alpha \in [0,1]$.
- $f : X \to \mathbb{R}$ is *strictly quasi-concave* if
  $f((1-\alpha)x + \alpha x') > f(x)$
  for all $x, x' \in A$ with $x \neq x'$ such that $f(x') \geq f(x)$ and all $\alpha \in (0,1)$.
- f is (strictly) quasi-convex if $-f$ is (strictly) quasi-concave.

Characterization via Gradient

Let $X \subset \mathbb{R}^N$ be a nonempty open convex set.

Proposition 5.22

Suppose that $f\colon X \to \mathbb{R}$ is differentiable.

1. *f is quasi-concave if and only if for all $x, x' \in X$,*

$$f(x') \geq f(x) \Rightarrow \nabla f(x) \cdot (x' - x) \geq 0. \tag{3}$$

2. *If f is quasi-concave, then for all $x, x' \in X$,*

$$f(x') > f(x),\ \nabla f(x) \neq 0 \Rightarrow \nabla f(x) \cdot (x' - x) > 0. \tag{4}$$

Proof

1. "Only if" part
   - Suppose that f is quasi-concave.
     Fix any $x, x' \in X$, and assume that $f(x') \geq f(x)$.
     Consider the function $g(t) = f((1-t)x + tx')$.
   - By quasi-concavity, $g(t) \geq g(0)$ for all $t \in [0,1]$.
   - Therefore, $g'(0) \geq 0$, where $g'(0) = \nabla f(x) \cdot (x' - x)$.

Proof

1. "If" part

   - Suppose that f is not quasi-concave.

     Then there exist $\bar{x}, \bar{x}' \in X$, $\bar{x} \neq \bar{x}'$, and $\bar{\alpha} \in [0,1]$ such that $f(\bar{x}') \geq f(\bar{x}) > f((1-\bar{\alpha})\bar{x} + \bar{\alpha}\bar{x}')$.

   - Consider the function $g(t) = f((1-t)\bar{x} + t\bar{x}')$.

   - Let $M = \min_{t\in[0,1]} g(t) < g(0)$, and let $\alpha^* = \min\{t \in [0,1] \mid g(t) = M\}$ (which is well defined by the continuity of g).

- By the continuity of g, there exists $\delta > 0$ such that $g(t) < g(0)$ for all $t \in (\alpha^* - \delta, \alpha^*)$.
- By the Mean Value Theorem, there exists $\alpha^{**} \in (\alpha^* - \delta, \alpha^*)$ such that $g'(\alpha^{**}) = \frac{g(\alpha^*) - g(\alpha^* - \delta)}{\delta} < 0$.
- Therefore, letting $x^{**} = (1 - \alpha^{**})\bar{x} + \alpha^{**}\bar{x}'$, we have

  $$g(0) = f(\bar{x}) > g(\alpha^{**}) = f(x^{**})$$

  and

  $$g'(\alpha^{**}) = \nabla f(x^{**}) \cdot (x^{**} - \bar{x}) < 0.$$

2.

- Suppose that f is quasi-concave and that $f(x') > f(x)$ and $\nabla f(x) \neq 0$.
- By the continuity of f, we have $f(x' - \varepsilon \nabla f(x)) > f(x)$ for some small $\varepsilon > 0$.
- Then by part 1, we have $\nabla f(x) \cdot ((x' - \varepsilon \nabla f(x)) - x) \geq 0$, or $\nabla f(x) \cdot (x' - x) \geq \varepsilon \|\nabla f(x)\|^2 > 0$.

Characterization via Gradient

Let $X \subset \mathbb{R}^N$ be a nonempty open convex set.

Proposition 5.23

Suppose that $f\colon X \to \mathbb{R}$ is differentiable.

1. *If for all $x, x' \in X$,*

$$f(x') \geq f(x),\ x \neq x' \Rightarrow \nabla f(x) \cdot (x' - x) > 0, \tag{5}$$

   then f is strictly quasi-concave.

2. *If f is strictly quasi-concave, then for all $x, x' \in X$,*

$$\begin{aligned} & f(x') \geq f(x),\ x \neq x',\ \nabla f(x) \neq 0 \\ & \Rightarrow \nabla f(x) \cdot (x' - x) > 0. \end{aligned} \tag{6}$$

Proof

1.

- Suppose that condition (5) holds.
- By part 1 of Proposition 5.22, f is quasi-concave.
- Assume that f is not strictly quasi-concave.

  Then there exist $\bar{x}, \bar{x}' \in X$, $\bar{x} \neq \bar{x}'$, and $\bar{\alpha} \in (0,1)$ such that $f(\bar{x}') \geq f(\bar{x}) \geq f(\bar{x}'')$, where $\bar{x}'' = (1-\bar{\alpha})\bar{x} + \bar{\alpha}\bar{x}'$ ($\neq \bar{x}, \bar{x}'$).
- Consider the function $g(t) = f((1-t)\bar{x} + t\bar{x}')$, which is quasi-concave.
- Since $g(0) \geq g(\bar{\alpha})$, by part 1 of Proposition 5.22 we have $g'(\bar{\alpha})(0-\bar{\alpha}) \geq 0$, or $g'(\bar{\alpha}) \leq 0$, where $g'(\bar{\alpha}) = \nabla f(\bar{x}'') \cdot (\bar{x}' - \bar{x}) = \frac{1}{1-\bar{\alpha}} \nabla f(\bar{x}'') \cdot (\bar{x}' - \bar{x}'')$.
- This contradicts condition (5) (with $x = \bar{x}''$ and $x' = \bar{x}'$).

2.

- Suppose that f is strictly quasi-concave and that $f(x') \geq f(x)$, $x \neq x'$, and $\nabla f(x) \neq 0$.
- By strict quasi-concavity, $f\left(\frac{1}{2}x + \frac{1}{2}x'\right) > f(x)$.
- Then by part 2 of Proposition 5.22, we have $\nabla f(x) \cdot \left(\left(\frac{1}{2}x + \frac{1}{2}x'\right) - x\right) > 0$, or $\frac{1}{2}\nabla f(x) \cdot (x' - x) > 0$.

Characterization via Hessian

Let $X \subset \mathbb{R}^N$ be a nonempty open convex set.

Proposition 5.24

Suppose that $f\colon X \to \mathbb{R}$ is differentiable and ∇f is differentiable.

- *If f is quasi-concave, then for all $x \in X$, $D^2 f(x)$ is negative semi-definite on $\{z \in \mathbb{R}^N \mid \nabla f(x) \cdot z = 0\}$, i.e.,*

  $$z \cdot D^2 f(x) z \leq 0$$

  for all $z \in \mathbb{R}^N$ with $\nabla f(x) \cdot z = 0$.

- *If for all $x \in X$, $D^2 f(x)$ is negative definite on $\{z \in \mathbb{R}^N \mid \nabla f(x) \cdot z = 0\}$, i.e.,*

  $$z \cdot D^2 f(x) z < 0$$

  for all $z \neq 0$ with $\nabla f(x) \cdot z = 0$, then f is strictly quasi-concave.