5. Differentiation

Takuma Habu*
takumahabu@cornell.edu

26th August 2024

1 Univariate functions

Definition 1. Suppose X is a linear space. A function $f : X \to \mathbb{R}$ is *univariate* if $\dim X = 1$ and *multivariate* if $\dim X > 1$.

Definition 2. A function $f : X \subseteq \mathbb{R} \to \mathbb{R}$ is *differentiable* at $x_0 \in \operatorname{int}(X)$ if the following limit exists:

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

When the limit exists, it is called the *derivative* of f at x_0, and we denote it by $f'(x_0)$ or $Df(x_0)$. Moreover, the function f is *differentiable* on $\operatorname{int}(S) \subseteq X$ if it is differentiable at all $x \in \operatorname{int}(S)$, and f is *differentiable* if it is differentiable on $\operatorname{int}(X)$.

Remark 1. We can think of the derivative of f at x_0, $f'(x_0)$, as an affine approximation of the function f at x_0. To see this, let's agree that the value of the approximating affine function, $\widehat{f}(x) = ax + b$ for some $a, b \in \mathbb{R}$, should coincide with the value of f at x_0; i.e.,

$$\widehat{f}(x_0) = ax_0 + b = f(x_0) \Leftrightarrow b = f(x_0) - ax_0.$$

Hence, we can rewrite $\widehat{f}$ as

$$\widehat{f}(x) = a(x - x_0) + f(x_0).$$

It remains to finding the value of a such that $\widehat{f}(x_0)$ "best" approximates f around x_0. One way is to find the slope of the line between x_0 and $x_0 + h$ for some $h > 0$ small. The slope is given by:

$$\frac{f(x_0 + h) - f(x_0)}{(x_0 + h) - x_0} = \frac{f(x_0 + h) - f(x_0)}{h.}.$$

If f is continuous from the "right" at x_0 (so that small increases from x_0 leads to small changes in $f(\cdot)$), the approximation should become better as $h \to 0$. This leads us to set a equal to

$$a^+ := \lim_{h \searrow 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

*Thanks to Giorgio Martini, Nadia Kotova and Suraj Malladi for sharing their lecture notes, on which these notes are heavily based.

where $h \searrow 0$ means for any decreasing sequence that converges to 0. Alternatively, thinking of f as being continuous from the "left" at x_0(so that small decreases from x_0 leads to small changes in $f(\cdot)$), we might set a equal to

$$a^- := \lim_{h \nearrow 0} \frac{f(x_0 - h) - f(x_0)}{h},$$

where $h \nearrow 0$ means any increasing sequence that converges to 0. Of course, without any assumption on f, these limits might not exist. However, when they exist, we call a^+ the *right-derivative* of f at x_0, and a^- the *left-derivative* of f at x_0. For these limits to exist, it must be the case that there is an open ball around x_0; i.e., x_0 must be contained in an open set. Moreover, if they both exist, then uniqueness of limit means that $a^+ = a^-$.

Now suppose that f has a derivative at x_0 then,

$$\widehat{f}(x) = f'(x_0)(x - x_0) + f(x_0).$$

Define the approximation error rate as

$$\epsilon(h) := \frac{f(x_0 + h) - \widehat{f}(x_0 + h)}{h}.$$

Then,

$$\lim_{h \to 0} \epsilon(h) = \lim_{h \to 0} \frac{f(x_0 + h) - f'(x_0)h - f(x_0)}{h} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} - f'(x_0) = 0.$$

In words, above says that the error rate, $\frac{\epsilon(h)}{h}$, between f and the approximating affine function $\widehat{f}$ vanishes in the limit $h = 0$.

Recalling the various definition of limits gives the following.

Proposition 1. *Suppose $f : X \subseteq \mathbb{R} \to \mathbb{R}$. The following are equivalent.*

(i) *f is differentiable at $x_0 \in \operatorname{int}(X)$.*

(ii) *for some $x \in \mathbb{R}$, $x = \lim_{h \to 0} \frac{f(x_0+h) - f(x_0)}{h}$.*

(iii) *for some $x \in \mathbb{R}$, $x = \lim_{n \to \infty} \frac{f(x_n) - f(x)}{x_n - x}$ for all sequences $(x_n)_n$ in X such that $x_n \to x_0$.*

(iv) *for some $x \in \mathbb{R}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that $|\frac{f(x_0+h) - f(x_0)}{h} - a| < \epsilon$ for all $h \in \mathbb{R}$ such that $x + h \in X$ and $|h| < \delta$.*

Proposition 2. *If $f : X \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable at $x_0 \in X$, then f is continuous at x_0.*

Proof. Observe that

$$\begin{aligned} 0 = f'(x_0) \cdot 0 &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \cdot \lim_{x \to x_0} (x - x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}(x - x_0) \\ &= \lim_{x \to x_0} f(x) - f(x_0). \end{aligned}$$

Hence, $\lim_{x \to x_0} f(x) = f(x_0)$. ■

Exercise 1 (PS7)**.** TFU: If $f : X \subseteq \mathbb{R} \to \mathbb{R}$ is continuous at $x_0 \in \text{int}(X)$, then f is differentiable at x_0.

Proposition 3. *Suppose $f, g : X \subseteq \mathbb{R} \to \mathbb{R}$ and that f and g are differentiable at $x_0 \in X$. Then,*

(i) $(f+g)'(x_0) = f'(x_0) + g'(x_0)$;

(ii) $(\alpha f)'(x_0) = \alpha f'(x_0)$ *for all* $\alpha \in \mathbb{R}$;

(iii) (product rule) $(f \cdot g)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$;

(iv) (quotient rule) $(\frac{f}{g})'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}$ *if* $g(x_0) \neq 0$.

Exercise 2. Prove Proposition 3.

Example 1. Suppose $f : X \subseteq \mathbb{R} \to \mathbb{R}$.

- ▷ f is a *constant function* if $f(\cdot) = c$ for some $c \in \mathbb{R}$. Then, $f'(x) = 0$ for all $x \in X$.
- ▷ f is a *monomial* if $f(\cdot) = (\cdot)^n$ for some $n \in \mathbb{N}$ for all $x \in X$. Then, $f'(x) = nx^{n-1}$ because

$$x^n - x_0^n = (x - x_0)\left(x^{n-1} + x^{n-2}x_0 + \cdots + xx_0^{n-2} + x_0^{n-1}\right).$$

  (In fact, this formula holds for any $n \in \mathbb{R}\backslash\{0\}$ but this is more difficult to show!)

- ▷ f is a *absolute value function* if $f(\cdot) = |\cdot|$. Then, $f'(x) = 1$ when $x > 0$ and $f'(x) = -1$ when $x < 0$. However, $f'(0)$ does not exist.

Proposition 4 (Chain Rule)**.** *Suppose $f : X \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable at $x_0 \in \text{int}(X)$ and that $g : Y \to \mathbb{R}$, where $f(X) \subseteq Y$, and g is differentiable at $f(x_0)$. Then, $g \circ f$ is differentiable at x_0 and*

$$(g \circ f)'(x_0) = (g' \circ f)(x_0) \cdot f'(x_0).$$

Exercise 3 (PS7)**.** Prove Proposition 4.

Theorem 1 (L'Hôpital's rule)**.** *Let $-\infty \leq a < b \leq +\infty$ and $f : (a,b) \to \mathbb{R}$ and $g : (a,b) \to \mathbb{R}\backslash\{0\}$ are differentiable on (a,b). If $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ are both 0 or $\pm\infty$, and $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ is finite value or is $\pm\infty$, then*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

The statement also holds for $x \to b$.

Example 2. Consider the constant relative risk aversion (CRRA) utility function, $u : \mathbb{R} \to \mathbb{R}$ such that

$$u(x,\gamma) \coloneqq \frac{x^{1-\gamma} - 1}{1-\gamma}.$$

Consider the limit of $u(x,\gamma)$ as $\gamma \to 1$. Note that $u(x, 1) =$ "0/0" and so we can apply L'Hôpital's rule to obtain

$$\lim_{\gamma\to 1} \frac{x^{1-\gamma} - 1}{1-\gamma} = \lim_{\gamma\to 1} \frac{e^{(1-\gamma)\ln(x)} - 1}{1-\gamma} = \lim_{\gamma\to 1} \frac{-\ln(x)\, e^{(1-\gamma)\ln x}}{-1} = \ln x.$$

Hence, CRRA is a generalisation of log utility!

Proposition 5. *Suppose $f : (a,b) \subseteq \mathbb{R} \to \mathbb{R}$ and f is strictly increasing and differentiable on (a,b). Then,*

$$\left(f^{-1}\right)'(f(x)) = \frac{1}{f'(x)} \ \forall x \in (a,b).$$

Exercise 4 (PS7)**.** Prove Proposition 5.[1]

Proposition 6. *Suppose $f : S \subseteq \mathbb{R} \to \mathbb{R}$. If $x_0 \in S$ is a local maximum or minimum of f and $f'(x)$ exists, then $f'(x) = 0$.*

Proof. Suppose x_0 is a local maximum of f (the other case when x_0 is a local minimum is analogous) and that $f'(x_0) = \lim_{x \to x_0} \frac{f(x_0)-f(x_0)}{x-x_0}$ exists. Let $\epsilon > 0$ be such that $f(x_0) \geq f(x)$ for all $x \in B_\epsilon(x_0)$. If $|x - x_0| < \epsilon$, then $f(x) - f(x_0) \leq 0$ is nonpositive. This implies that for x sufficiently close to x_0, $\frac{f(x)-f(x_0)}{x-x_0} \geq 0$ if $x < x_0$ and $\frac{f(x)-f(x_0)}{x-x_0} \leq 0$ if $x > x_0$. Then, because $f'(x_0)$ exists, the limit of $\frac{f(x)-f(x_0)}{x-x_0}$ as $x \searrow 0$ and $x \nearrow 0$ must agree; i.e., $f'(x_0) = 0$. ■

Theorem 2 (Rolle's Theorem)**.** *Let $[a,b]$ be a closed and bounded interval in $\mathbb{R}$ and suppose $f : [a,b] \to \mathbb{R}$ is continuous. If f is differentiable on (a,b) and $f(a) = f(b)$, then $f'(c) = 0$ for some $c \in (a,b)$.*

Proof. Since f is continuous and $[a,b]$ is compact, Weierstrass Extreme Value Theorem tells us that there exists $u, \ell \in [a,b]$ such that

$$f(\ell) \leq f(x) \leq f(u) \ \ \forall x \in [a,b].$$

Suppose f is differentiable on (a,b) and $f(a) = f(b)$. If $\{\ell, u\} \subseteq \{a, b\}$, then f must be a constant function (why?), and so $f'(x) = 0$ for all $x \in (a,b)$. If this is not the case, then either $\ell \in (a,b)$ or $u \in (a,b)$. In the former case, because f has a minimum at ℓ, we must have $f'(\ell) = 0$ by Proposition 6. In the latter case, f has a maximum at u, and so $f'(u) = 0$ by Proposition 6 again. ■

Corollary 1 (Mean Value theorem)**.** *Let $[a,b]$ be a closed and bounded interval in $\mathbb{R}$ and suppose $f : [a,b] \to \mathbb{R}$ is continuous and differentiable on (a,b). Then, there exists $x \in (a,b)$ such that*

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

Proof. Define

$$g(t) \coloneqq f(t) - f(a) + \frac{a-t}{b-a}(f(b) - f(a)).$$

Observe that $g : [a,b] \to \mathbb{R}$ is continuous, $g(a) = 0 = g(b)$. Hence, by Rolle's Theorem, there exists $x \in (a,b)$ such that

$$g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a} = 0 \Leftrightarrow f(b) - f(a) = f'(x)(b-a).$$ ■

Remark 2. Let $a = x_0$ and $b = x_0 + h$. The mean value theorem tells us that, for some $x \in (x_0, x_0 + h)$,

$$f(x_0 + h) = f(x_0) + f'(x)h.$$

[1]A heuristic proof is to differentiate the following identity while using chain rule:

$$f^{-1}(f(x)) \equiv x \Rightarrow \left(f^{-1}\right)' f(x) f'(x) = 1 \Rightarrow \left(f^{-1}\right)' f(x) = \frac{1}{f'(x)}.$$

Thus, the theorem gives us a way of approximating f around x_0 via an affine function $f(x_0)+f'(x)h$.

Corollary 2. *Let $[a,b]$ be an a closed and bounded interval in $\mathbb{R}$ and suppose $f:[a,b]\to\mathbb{R}$ is continuous and differentiable on (a,b).*

(i) If $f'(x)\geq 0$ for all $x\in(a,b)$, then f is monotonically increasing.

(ii) If $f'(x)=0$ for all $x\in(a,b)$, then f is constant.

(iii) If $f'(x)\leq 0$ for all $x\in(a,b)$, then f is monotonically decreasing.

Exercise 5 (PS7)**.** Prove part Corollary 2.

2 Taylor expansions

Suppose f has a derivative f' on some interval $(a,b)\subseteq\mathbb{R}$ and that f' is itself differentiable on (a,b). Then, the derivative of f' is denoted f'' or D^2f. Similarly, if f'' is differentiable on $f'((a,b))$, then the derivative of f'' is denoted f''' or D^3f (and so on).

Given an interval $[a,b]\subseteq\mathbb{R}$, let $\mathbf{C}[a,b]$ denote the set of all continuous, real-valued function defined on $[a,b]$. A function $f\in\mathbf{C}[a,b]$ is *continuously differentiable* if f' is a continuous function on $[a,b]$; i.e., $f'\in\mathbf{C}[a,b]$. The set of all such functions is denoted $\mathbf{C}^1[a,b]$; i.e.,

$$\mathbf{C}^1[a,b]\coloneqq\{f\in\mathbf{C}[a,b]:f'\in\mathbf{C}[a,b]\}.$$

More generally, for any $k\in\mathbb{N}$, the set of all *k-times continuously differentiable* functions is defined inductively as

$$\mathbf{C}^k[a,b]\coloneqq\left\{f\in\mathbf{C}[a,b]:f'\in\mathbf{C}^{k-1}[a,b]\right\},$$

where $\mathbf{C}^0\coloneqq\mathbf{C}$. Thus, if $f\in\mathbf{C}^k[a,b]$, then f has up to and including the kth derivative, and f and all its derivatives up to and including the kth derivative are all continuous on $[a,b]$. Finally, f is *smooth* if $f\in\mathbf{C}^k[a,b]$ *for all* $k\in\mathbb{N}$ *and* $\mathbf{C}^\infty[a,b]$ denotes the set of smooth functions.

Theorem 3 (Taylor Theorem)**.** *Suppose $f\in\mathbf{C}^k[a,b]$ and let α and β be two distinct points in $[a,b]$. Define*

$$\begin{aligned}P_{k-1}(x)&\coloneqq\sum_{n=1}^{k}\frac{f^{(n-1)}(\alpha)}{(n-1)!}(x-\alpha)^{n-1}\\&=f(\alpha)+f'(\alpha)(x-\alpha)+\frac{1}{2}f''(\alpha)(x-\alpha)^2+\cdots+\frac{f^{(k-1)}(\alpha)}{(k-1)!}(x-\alpha)^{k-1}.\end{aligned}$$

Then, there exists x between α and β such that

$$f(\beta)=P_{k-1}(\beta)+\frac{f^{(k)}(x)}{k!}(\beta-\alpha)^k.$$

Proof. Define E as the error rate of approximating $f(\beta)$ via just $P_{k-1}(\beta)$; i.e.,

$$E\coloneqq\frac{f(\beta)-P_{k-1}(\beta)}{(\beta-\alpha)^k}\Leftrightarrow f(\beta)=P_{k-1}(\beta)+E(\beta-\alpha)^k.$$

Define g as

$$g(x) \coloneqq f(x) - P_{k-1}(x) - E(x-\alpha)^k.$$

Because $P_{k-1}^{(n)}(\alpha) = f^{(n)}(\alpha)$ for all $n \in \{1, \ldots, k\}$, we have $g^{(n)}(\alpha) = 0$ for all $n \in \{1, \ldots, k\}$. We also have

$$g(\alpha) = 0 = g(\beta).$$

By Rolle's theorem, there exists x_1 between α and β such that $g'(x_1) = 0$. Since $g'(\alpha) = 0 = g'(x_1)$, by Rolle's theorem again, there exists x_2 between α and x_1 such that $g''(x_2) = 0$. Because x_1 lies between α and β, x_2 must also lie between α and β. Continuing in this way, there exists x_k between α and β such that $g^{(k)}(x_k) = 0$. Since

$$P_{k-1}^{(k)}(x) = 0, \ D^k(x-\alpha)^k = k!,$$

we have

$$g^{(k)}(x) = f^{(k)}(x) - P_{k-1}^{(k)}(x) - E \cdot D^k(x-\alpha)^k = f^{(k)}(x) - E(k!)$$

so that

$$g^{(k)}(x_k) = 0 = f^{(k)}(x_k) - E(k!) \Rightarrow E = \frac{f^{(k)}(x_k)}{k!}.$$

Since

$$\begin{aligned} 0 &= g(\beta) = f(\beta) - P_{k-1}(\beta) - E(\beta-\alpha)^k \\ \Rightarrow f(\beta) &= P_{k-1}(\beta) + E(\beta-\alpha)^k = P_{k-1}(\beta) + \frac{f^{(k)}(x_k)}{k!}(\beta-\alpha)^k. \end{aligned}$$ ■

Remark 3. If $k = 1$, above says that $P_0(x) = P(\alpha)$ and so $f(\beta) = f(\alpha) + f'(x)(\beta-\alpha)$ for some $x \in (\alpha, \beta)$; i.e., above reduces to the Mean Value theorem. Taylor theorem generalises the affine approximation of f: it tells us that we can approximate f using polynomials of degree $k-1$, where $P_{k-1}(\beta)$ is the approximating function and the term $\frac{f^k(x)(\beta-\alpha)^k}{k!}$ is the error term.

The reason why we obtain polynomial can be seen as follows: we know that $f(x_0+h) = f(x_0) + f'(x_1)h$ for some $x_1 \in (x_0, x_0+h)$. Say we approximate $f(x_0+h)$ as $f(x_0) + f'(x_0+h)h$. Since f' is also continuous and differentiable, we can similarly approximate $f'(x_0+h) = f'(x_0) + f''(x_0)h$. If we combine them, we obtain

$$f(x_0+h) \approx f(x_0) + (f'(x_0) + f''(x_0)h)h = f(x_0) + f'(x_0)h + f''(x_0)h^2,$$

which gives us a polynomial of degree 2.

Remark 4. Suppose we define $\alpha \coloneqq x_0$ and $\beta \coloneqq \alpha + h$, then we can rewrite above as

$$\begin{aligned} f(\alpha+h) &= P_{k-1}(\alpha+h) + \frac{f^k(x)}{k!}h^k, \\ P_{k-1}(\alpha+h) &= f(h) + f'(\alpha)h + \frac{1}{2}f''(\alpha)h^2 + \cdots + \frac{f^{(k-1)}(\alpha)}{(k-1)!}h^{k-1} \end{aligned}$$

for some $x \in (\alpha, \alpha + h)$. If we further have that $f \in \mathbf{C}^k[a,b]$ so that $f^{(k)}$ is continuous at α, then

$$\lim_{h \to 0} \frac{\frac{f^{(k)}(x)}{k!} h^k}{h^{k-1}} = \lim_{h \to 0} \frac{f^{(k)}(x)}{k!} h = \frac{f^{(k)}(\alpha)}{k!} \lim_{h \to 0} h = 0;$$

i.e., this means that the error term $\frac{f^k(x)}{k!} h^k$ converges to zero at a rate faster than h^{k-1}; i.e., the error term, which is the kth power of h, is *small* relative to the $k-1$-degree polynomial approximation of $f(\beta)$, P_{k-1}. We therefore often approximate functions using polynomials using $(k-1)$*th-order Taylor expansion*; i.e.,

$$\tilde{f}(\alpha + h) \approx P_{k-1}(\alpha + h) = \sum_{n=0}^{k-1} \frac{f^{(n)}(\alpha)}{(n-1)!} h^n,$$

where $f^{(0)} := f$.

Remark 5. *Maclaurin series* is an infinite Taylor expansion of functions around 0. Here are some useful ones to know:

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \qquad (1) \\ \ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} x^n}{n} \ \forall x \in (-1, 1] \\ \ln(1-x) &= -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots = \sum_{n=1}^{\infty} -\frac{x^n}{n} \ \forall x \in (-1, 1]. \end{aligned}$$

Corollary 3. *Suppose* $f : (a,b) \subseteq \mathbb{R} \to \mathbb{R}$, $f \in \mathbf{C}^k$ *and that* $f'(x_0) = f''(x_0) = \ldots = f^{(k-1)}(x_0) = 0$ *and* $f^{(k)}(x_0) \neq 0$. *Then,*

(i) if k is even and $f^{(k)}(x_0) > 0$, then f has a local minimum at x_0;

(ii) if k is even and $f^{(k)}(x_0) < 0$, then f has a local maximum at x_0;

(iii) if k is odd, then f has neither a local minima or a local minima (i.e., x_0 is an inflection point).

Exercise 6 (PS7)**.** Prove Corollary 3. **Hint:** If g is continuous and $g(x_0) > 0$, then $g > 0$ in some neighbourhood of x_0.

3 Multivariate functions

Definition 3. Let $V \subseteq \mathbb{R}^d$ and $W \subseteq \mathbb{R}^m$ be two linear spaces. A function $f : V \to W$ is a *linear transformation* if

$$f(\alpha \mathbf{x} + \mathbf{y}) = \alpha f(\mathbf{x}) + f(\mathbf{y}) \ \forall \mathbf{x}, \mathbf{y} \in V, \ \forall \alpha \in \mathbb{R}.$$

Remark 6. A linear transformation preserves linear combinations in one space to another.

Definition 4. The function $f : X \subseteq \mathbb{R}^d \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0 \in \text{int}(X)$, if there exists some linear transformation $A : \mathbb{R}^d \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - (f(\mathbf{x}_0) + A(\mathbf{h}))\|_m}{\|\mathbf{h}\|_d} = 0.$$

When it exists, the linear transformation A is called the *derivative* of f at $\mathbf{x}_0$, and we denote it as $f'(\mathbf{x}_0)$ or $Df(\mathbf{x}_0)$. Moreover, the function f is *differentiable* on $\text{int}(S) \subseteq X$ if it is differentiable at all $x \in \text{int}(S)$, and f is *differentiable* if it is differentiable on $\text{int}(X)$.

Remark 7. To see how this relates to the case when $d = m = 1$, recall that, if $f : \mathbb{R} \to \mathbb{R}$ has a derivative $a \in \mathbb{R}$ at x_0, then

$$a = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \Leftrightarrow 0 = \lim_{h \to 0} \frac{f(x+h) - (f(x) + ah)}{h}.$$

Above means that, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\left| \frac{f(x+h) - (f(x) + ah)}{h} \right| = \frac{|f(x+h) - (f(x) + ah)|}{|h|} < \epsilon \ \forall h > 0 : |h| < \delta;$$

i.e., the derivative of f at x_0, a, satisfies

$$\lim_{h \to 0} \frac{|f(x+h) - (f(x) + ah)|}{|h|} = 0.$$

Thinking of ah as a linear transformation between the domain of f and the codomain of f gives the definition of total derivative above.

Proposition 7. *The derivative of $f : \mathbb{R}^d \to \mathbb{R}^m$ is unique whenever it exists.*

Proof. Suppose that A_1 and A_2 both satisfy the definition of a derivative at some point $\mathbf{x}_0$. By linearity, for any $\mathbf{h} \in \mathbb{R}^m$,

$$\begin{aligned}
&\|A_1(\mathbf{h}) - A_2(\mathbf{h})\| \\
&= \|[f(\mathbf{x}_0 + \mathbf{h}) - (f(\mathbf{x}_0) + A_2(\mathbf{h}))] - [f(\mathbf{x}_0 + \mathbf{h}) - (f(\mathbf{x}_0) + A_1(\mathbf{h}))]\| \\
&\leq \|[f(\mathbf{x}_0 + \mathbf{h}) - (f(\mathbf{x}_0) + A_2(\mathbf{h}))]\| + \|[f(\mathbf{x}_0 + \mathbf{h}) - (f(\mathbf{x}_0) + A_1(\mathbf{h}))]\|.
\end{aligned}$$

Since $\| \cdot \| \geq 0$ and by the definition of A_1 and A_2 as derivatives of f at $\mathbf{x}_0$:

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|A_1(\mathbf{h}) - A_2(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

So in particular, for any nonzero $\mathbf{h}' \in \mathbb{R}^d$ (i.e., $\|\mathbf{h}'\| > 0$),

$$\begin{aligned}
0 &= \lim_{t \to 0} \frac{\|A_1(t\mathbf{h}') - A_2(t\mathbf{h}')\|}{\|t\mathbf{h}'\|} = \lim_{t \to 0} \frac{|t| \, \|A_1(\mathbf{h}') - A_2(\mathbf{h}')\|}{|t| \, \|\mathbf{h}'\|} = \lim_{t \to 0} \frac{\|A_1(\mathbf{h}') - A_2(\mathbf{h}')\|}{\|\mathbf{h}'\|} \\
&= \frac{\|A_1(\mathbf{h}') - A_2(\mathbf{h}')\|}{\|\mathbf{h}'\|}.
\end{aligned}$$

Hence, $\|A_1(\mathbf{h}') - A_2(\mathbf{h}')\| = 0$ for any nonzero $\mathbf{h}'$, which implies that A_1 and A_2 are equivalent. ∎

Remark 8. Any linear transformation from $\mathbb{R}^d$ to $\mathbb{R}^m$ corresponds to an $m \times d$ matrix.[2] We therefore write a $A(\mathbf{h}) \equiv A\mathbf{h}$, where $\mathbf{h}$ is treated as a $d \times 1$ column vector. The derivative of $f : \mathbb{R}^d \to \mathbb{R}^m$

[2] Formally, this follows from the fact that the space of linear transformation between finite-dimensional linear spaces are isomorphic to matrices and we obtain the unique matrix that corresponds to a linear transformation using the standard bases for the linear spaces $\mathbb{R}^d$ and $\mathbb{R}^m$.

at $\mathbf{x}_0$, $Df(\mathbf{x}_0)$, is an $m \times d$ matrix that we often refer to as the *total derivative of* f *at* $\mathbf{x}_0$. *When* $m = 1$, Df is a $1 \times d$ row vector and the derivative is called the *gradient* of f at $\mathbf{x}_0$, denoted $\nabla f(\mathbf{x}_0)$, where

$$\nabla f(\mathbf{x}_0) \equiv \begin{bmatrix} \frac{\mathrm{d}f}{\mathrm{d}x_1}(\mathbf{x}_0) & \cdots & \frac{\mathrm{d}f}{\mathrm{d}x_d}(\mathbf{x}_0) \end{bmatrix}_{1 \times d}.$$

Since a multivariate vector function $f : \mathbb{R}^d \to \mathbb{R}^m$ can be written as an ordered collection of real-valued functions $(f_i : \mathbb{R}^n \to \mathbb{R})_{i=1}^m$, in particular, we can write $f(\mathbf{x})$ as a column vector:

$$f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}_{m \times 1}.$$

Since the derivatives are just limits, the following result that allows us to extend many of the results for multivariate functions (i.e., $f : \mathbb{R}^n \to \mathbb{R}$) into functions from multivariate vector functions (i.e., $f : \mathbb{R}^n \to \mathbb{R}^m$) is immediate.

Proposition 8. *Suppose* $f : X \subseteq \mathbb{R}^d \to \mathbb{R}^m$. *Then,* f *is differentiable at* $\mathbf{x}_0 \in \operatorname{int}(X)$ *if and only if each corresponding* f_i *is differentiable at* $\mathbf{x}_0$. *Moreover,*

$$Df(\mathbf{x}_0) = \begin{bmatrix} \nabla f_1(\mathbf{x}_0) \\ \vdots \\ \nabla f_m(\mathbf{x}_0) \end{bmatrix}_{m \times d} \equiv \begin{bmatrix} \frac{\mathrm{d}f_1}{\mathrm{d}x_1}(\mathbf{x}_0) & \cdots & \frac{\mathrm{d}f_1}{\mathrm{d}x_d}(\mathbf{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\mathrm{d}f_m}{\mathrm{d}x_1}(\mathbf{x}_0) & \cdots & \frac{\mathrm{d}f_m}{\mathrm{d}x_d}(\mathbf{x}_0) \end{bmatrix}.$$

Remark 9. We can still view derivatives as an approximation by an affine function of the form $f(\mathbf{x}) + Df\mathbf{h}$, where

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + Df\mathbf{h} + \epsilon(\|\mathbf{h}\|),$$

where $\epsilon(\|\mathbf{h}\|)$ is an error term such that

$$\lim_{\mathbf{h} \to 0} \frac{\epsilon(\|\mathbf{h}\|)}{\|\mathbf{h}\|} = 0.$$

Uniqueness of Df implies that, Df is the bear linear approximation of f near $\mathbf{x}_0$.

Exercise 7. Prove that if $f : X \subseteq \mathbb{R}^d \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0 \in \operatorname{int}(X)$, then it is continuous at $\mathbf{x}_0$.

Proposition 9. *Suppose* $f, g : X \subseteq \mathbb{R}^d \to \mathbb{R}^m$ *are both differentiable on* $\operatorname{int}(X)$.

(i) $D(f + g) = Df + Dg$.

(ii) $D(\lambda f) = \lambda Df \ \forall \lambda \in \mathbb{R}$.

(iii) (product rule) if $m = 1$, $\nabla(f \cdot g) = \nabla f \cdot g + f \cdot \nabla g$ *($\cdot$ is a scalar multiplication).*

(iv) (quotient rule) if $m = 1, \nabla\left(\frac{f}{g}\right) = \frac{\nabla f \cdot g - f \cdot \nabla g}{g^2}$ *(at* $\mathbf{x} \in \operatorname{int}(X)$ *such that* $g(x) \neq 0$*).*

Proposition 10 (Chain Rule)**.** *Let* $X \subseteq \mathbb{R}^d$ *be open and suppose* $f : X \to \mathbb{R}^m$. *Let* $Y \subseteq f(S)$ *be open and suppose* $g : Y \to \mathbb{R}^k$. *If* f *is differentiable at* $\mathbf{x}_0 \in X$ *and* g *is differentiable at* $f(x)$, *then*

$g \circ f : X \to \mathbb{R}^k$ is differentiable at $\mathbf{x}_0$, and

$$D\left(g \circ f\right)\left(\mathbf{x}_0\right) = Dg\left(f\left(\mathbf{x}\right)\right) Df\left(\mathbf{x}_0\right),$$

where $Dg(f(\mathbf{x}_0)) \in \mathbb{R}^{k \times m}$, $Df(\mathbf{x}_0) \in \mathbb{R}^{m \times d}$ and $D(g \circ f)(\mathbf{x}_0) \in \mathbb{R}^{k \times d}$.

Theorem 4 (Mean Value Theorem)**.** *Suppose $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$ is differentiable on $\operatorname{int}(X)$ and $\ell(\mathbf{x}, \mathbf{y}) \subseteq \operatorname{int}(X)$, where $\ell(\mathbf{x}, \mathbf{y})$ is a line segment from $\mathbf{x}$ to $\mathbf{y}$.[3] Then, there exists $\mathbf{z} \in \ell(\mathbf{x}, \mathbf{y})$ such that*

$$f\left(\mathbf{x}\right) - f\left(\mathbf{y}\right) = \nabla f\left(\mathbf{z}\right)\left(\mathbf{x} - \mathbf{y}\right).$$

Remark 10. For functions whose codomain is $\mathbb{R}^m$, applying the theorem above gives us $\mathbf{z}_1, \ldots, \mathbf{z}_m \in \ell(\mathbf{x}, \mathbf{y})$ such that

$$f_i\left(\mathbf{y}\right) - f_i\left(\mathbf{x}\right) = Df_i\left(\mathbf{z}_i\right)\left(\mathbf{y} - \mathbf{x}\right) \ \forall i \in \left\{1, \ldots, m\right\}.$$

However, this does not guarantee that there is a single $\mathbf{z}$ that works for every component.

Looking at the definition of a total derivative, we see that it is the limit as $\mathbf{h} \to \mathbf{0}$; i.e., it is the limit of *any* sequence $(\mathbf{x}_n)_n$ in $\mathbb{R}^d$ that converges to $\mathbf{0}$. If $n = 3$, all sequences below are permissible:

$$\left(\left(\frac{1}{k}, \frac{1}{k}, \frac{1}{k}\right)\right)_{k \in \mathbb{N}}, \ \left(\left(0, \frac{1}{k}, 0\right)\right)_{k \in \mathbb{N}}, \ \left(\left(0, 0, \frac{1}{k}\right)\right)_{k \in \mathbb{N}}.$$

In the second sequence, only the second coordinate is converging while the other coordinates are always zero. In contrast, in the third sequence, only the last coordinate is converging while the other coordinates are always zero. You might wonder if the function f might behave differently across these different sequences. In other words, if we perturb only one coordinate at a time, how does the function f behave?

Definition 5. Given $f : X \subseteq \mathbb{R}^d \to \mathbb{R}^m$, the jth *partial derivative* of f_i at $\mathbf{x}_0 \in \operatorname{int}(X)$, if it exists, is defined as

$$\frac{\partial f_i}{\partial x_j}\left(\mathbf{x}_0\right) \coloneqq \lim_{h \to 0} \frac{f_i\left(\mathbf{x}_0 + h\mathbf{e}_j\right) - f_i\left(\mathbf{x}_0\right)}{h} \ \forall \left(i, j\right) \in \left\{1, \ldots, m\right\} \times \left\{1, \ldots, d\right\},$$

where $\mathbf{e}_j \in \mathbb{R}^d$ is the jth standard basis of $\mathbb{R}^d$. Observe that ith partial derivatives of f_i considers how the value of f_i changes when $\mathbf{x}$ moves in the direction of the ith coordinate.

Remark 11. The partial derivative $(\partial f_i / \partial x_j)(\mathbf{x}_0)$ is equivalent to computing the derivative of the univariate function, $\tilde{f}_i(h) \coloneqq f_i(\mathbf{x}_0 + h\mathbf{e}_j)$, and evaluating this at $h = 0$; i.e.,

$$\frac{\partial f_i}{\partial x_j}\left(\mathbf{x}_0\right) = D\tilde{f}_i\left(0\right).$$

The following establishes the connection between total and partial derivatives.

Proposition 11. *Suppose $f : X \subseteq \mathbb{R}^d \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0 \in \operatorname{int}(X)$. Then, $\frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)$ exists*

[3] That is, $\ell(\mathbf{x}, \mathbf{y}) \coloneqq \{\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} : \alpha \in [0, 1]\}$.

for any $(i,j) \in \{1,\dots,m\} \times \{1,\dots,d\}$ *and*

$$Df(\mathbf{x}_0) = \left[\frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)\right]_{ij} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \dots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{bmatrix}_{m \times d}.$$

Exercise 8 (PS8)**.** Prove Proposition 11.

Remark 12. The matrix corresponding of partial derivatives is called *Jacobian matrix* at $\mathbf{x}_0$. Thus, the result above tells us that the when f is differentiable, then its total derivative is given by the Jacobian matrix. It also tells us that with $m = 1$,

$$\frac{\mathrm{d}f}{\mathrm{d}x_i}(\mathbf{x}_0) = \frac{\partial f}{\partial x_i}(\mathbf{x}_0) \ \ \forall i \in \{1,\dots,n\}. \tag{2}$$

Exercise 9 (PS8)**.** Let $f(x,y) = \frac{xy}{x^2+y^2}$, if $(x,y) \neq (0,0)$, and let $f(0,0) = 0$. Show that the partial derivatives of f exist at $(0,0)$, but that f is not differentiable at $(0,0)$.

Remark 13. Proposition 11 says that if a function is differentiable, its partial derivatives exist. Moreover, Exercise Exercise 9 shows the converse is not necessarily true. The following says that the converse holds if the partial derivatives exist and are continuous.

Proposition 12. *Suppose* $f : X \subseteq \mathbb{R}^d \to \mathbb{R}^m$. *Then,* f *is differentiable at* $\mathbf{x}_0 \in \text{int}(X)$ *and* Df *is continuous at* $\mathbf{x}_0$ *if and only if* $\frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)$ *exists on an open ball around* $\mathbf{x}_0$ *and is continuous at* $\mathbf{x}_0$ *for all* $(i,j) \in \{1,\dots,m\} \times \{1,\dots,d\}$.

Proof. Fix $f : X \subseteq \mathbb{R}^d \to \mathbb{R}^m$ and $\mathbf{x} \in \text{int}(X)$. Suppose that all partial derivatives exists on an open ball around $\mathbf{x}_0$ and that they are also continuous at $\mathbf{x}_0$. We will prove the result for the case when $d = 2$ and $m = 1$. We will show first that $A = (\frac{\partial f}{\partial x_1}(\mathbf{x}_0), \frac{\partial f}{\partial x_2}(\mathbf{x}_0)) \in \mathbb{R}_{1\times 2}$ is the derivative of f at $\mathbf{x}_0$ meaning that f is differentiable at $\mathbf{x}_0$. That is, we wish to show that, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\frac{|f(x_{0,1}+h_1, x_{0,2}+h_2) - (f(\mathbf{x}_0) + A\mathbf{h})|}{\|\mathbf{h}\|} < \epsilon \ \forall \mathbf{h} \in \mathbb{R}^2 \setminus \{\mathbf{0}\} : \|\mathbf{h}\| < \delta.$$

Fix some $\epsilon > 0$ and any $\mathbf{h} = (h_1, h_2) \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$, letting $[x_{0,1}, x_{0,1} + h_1]$ and applying the mean value theorem to $f(\cdot, x_{0,2})$ gives

$$f(x_{0,1}+h_1, x_{0,2}) - f(\mathbf{x}_0) = \frac{\partial f}{\partial x_1}(x_1, x_{0,2})\, h_1$$

for some $x_1 \in (x_{0,1}, x_{0,1} + h_1)$. Mean value theorem also gives that, for some $x_2 \in (x_{0,2}, x_{0,2} + h_2)$,

$$f(x_{0,1}+h_1, x_{0,2}+h_2) - f(x_{0,1}+h_1, x_{0,2}) = \frac{\partial f}{\partial x_2}(x_{0,1}+h_1, x_2)\, h_2.$$

Then,

$$
\begin{aligned}
&|f(x_{0,1}+h_1, x_{0,2}+h_2) - (f(\mathbf{x}_0) + A\mathbf{h})| \\
&= \left| f(x_{0,1}+h_1, x_{0,2}+h_2) - \left(f(\mathbf{x}_0) + \frac{\partial f}{\partial x_1}(\mathbf{x}_0) h_1 + \frac{\partial f}{\partial x_2}(\mathbf{x}_0) h_2 \right) \right| \\
&= \left| \underbrace{f(x_{0,1}+h_1, x_{0,2}+h_2) - f(x_{0,1}+h_1, x_{0,2})}_{=\frac{\partial f}{\partial x_2}(x_{0,1}+h_1, x_2) h_2} + \underbrace{f(x_{0,1}+h_1, x_{0,2}) - f(\mathbf{x}_0)}_{=\frac{\partial f}{\partial x_1}(x_1, x_{0,2}) h_1} - \frac{\partial f}{\partial x_1}(\mathbf{x}_0) h_1 - \frac{\partial f}{\partial x_2}(\mathbf{x}_0) h_2 \right| \\
&= \left| \left[\frac{\partial f}{\partial x_2}(x_{0,1}+h_1, x_2) - \frac{\partial f}{\partial x_2}(\mathbf{x}_0) \right] h_2 + \left[\frac{\partial f}{\partial x_1}(x_1, x_{0,2}) - \frac{\partial f}{\partial x_1}(\mathbf{x}_0) \right] h_1 \right| \\
&\leq \left| \left[\frac{\partial f}{\partial x_2}(x_{0,1}+h_1, x_2) - \frac{\partial f}{\partial x_2}(\mathbf{x}_0) \right] h_2 \right| + \left| \left[\frac{\partial f}{\partial x_1}(x_1, x_{0,2}) - \frac{\partial f}{\partial x_1}(\mathbf{x}_0) \right] h_1 \right| \\
&\leq \left| \frac{\partial f}{\partial x_2}(x_{0,1}+h_1, x_2) - \frac{\partial f}{\partial x_2}(\mathbf{x}_0) \right| |h_2| + \left| \frac{\partial f}{\partial x_1}(x_1, x_{0,2}) - \frac{\partial f}{\partial x_1}(\mathbf{x}_0) \right| |h_1|
\end{aligned}
$$

(check that you know where the inequality come from). Since the partial derivatives are continuous at $\mathbf{x}_0$, there is a $\delta > 0$ such that if $\|\mathbf{h}\| < \delta$, then

$$\left| \frac{\partial f}{\partial x_2}(x_{0,1}+h_1, x_2) - \frac{\partial f}{\partial x_2}(\mathbf{x}_0) \right|, \left| \frac{\partial f}{\partial x_1}(x_1, x_{0,2}) - \frac{\partial f}{\partial x_1}(\mathbf{x}_0) \right| < \frac{\epsilon}{2}.$$

Thus, for $\|\mathbf{h}\| < \delta$,

$$|f(x_{0,1}+h_1, x_{0,2}+h_2) - (f(\mathbf{x}_0) + A\mathbf{h})| \leq \epsilon \|\mathbf{h}\|.$$

Hence, $Df = A$. That Df is continuous at $\mathbf{x}_0$ follows from the fact that the partial derivatives are all continuous at $\mathbf{x}_0$. ■

Exercise 10 (PS8)**.** Let f be a differentiable function from $(a, b) \subset \mathbb{R}$ into an open subset $Y \subset \mathbb{R}^d$. Let $g : Y \to \mathbb{R}$ be differentiable at $f(x_0)$ for $x_0 \in (a, b)$. Express $D(g \circ f)$ in terms of the partial derivatives of f and g.

Definition 6. Let $X \subseteq \mathbb{R}$ be open and consider $f : X \to \mathbb{R}^m$. If f is differentiable on open subset $X_1 \subseteq X$, then $Df : X_1 \to \mathbb{R}^{m \times d}$. In turn, if Df is differentiable on open subset $X_2 \subseteq X_1$, then $D^2 f : X_2 \to \mathbb{R}^{mn^2}$. If $D^2 f$ exists, f is said to be *twice differentiable*. More generally, the kth-order derivative of f at $\mathbf{x}_0 \in \text{int}(X)$ is an $mn^{k-1} \times n$ real matrix and is denoted $D^k(\mathbf{x}_0)$.

Definition 7. Let $X \subseteq \mathbb{R}$ be open. The function $f : X \to \mathbb{R}^m$ is *kth continuously differentiable at* $\mathbf{x}_0$, denoted $f \in \mathbf{C}^k$ *at* $\mathbf{x}_0$*,* if $\mathbf{x}_0 \in X$ and $D^k f(\mathbf{x}_0)$ is continuous at $\mathbf{x}_0$ (where X_k is the set of points at which $D^{k-1} f$ is differentiable). The function $f : X \to \mathbb{R}^m$ is *kth continuously differentiable*, denoted $f \in \mathbf{C}^k$, if f is kth continuously differentiable at all $\mathbf{x}_0 \in X$.

The second derivative of a real-valued function $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$ at $\mathbf{x}_0 \in \text{int}(X)$ is called the *Hessian matrix* of f at $\mathbf{x}_0$:

$$H_f(\mathbf{x}_0) \coloneqq D^2(\mathbf{x}_0) = \begin{bmatrix} \left(\nabla \frac{\partial f}{\partial x_1}\right)(\mathbf{x}_0) \\ \vdots \\ \left(\nabla \frac{\partial f}{\partial x_d}\right)(\mathbf{x}_0) \end{bmatrix} = \begin{bmatrix} \frac{\partial\left(\frac{\partial f}{\partial x_1}\right)}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial\left(\frac{\partial f}{\partial x_1}\right)}{\partial x_d}(\mathbf{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial\left(\frac{\partial f}{\partial x_n}\right)}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial\left(\frac{\partial f}{\partial x_d}\right)}{\partial x_d}(\mathbf{x}_0) \end{bmatrix}_{d \times d},$$

where $(\nabla \frac{\partial f}{\partial x_i})(\mathbf{x}_0)$ is the gradient of the function $\frac{\partial f}{\partial x_i}$ at $\mathbf{x}_0$ and the $\frac{\partial(\frac{\partial f}{\partial x_i})}{\partial x_j}(\mathbf{x}_0)$ is the jth partial derivative of the function $\frac{\partial f}{\partial x_i}$ at $\mathbf{x}_0$. We refer to the latter as a *cross partial* of f at $\mathbf{x}_0$ and write

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}_0) \coloneqq \frac{\partial\left(\frac{\partial f}{\partial x_i}\right)}{\partial x_j}(\mathbf{x}_0).$$

The following tells us that when f is twice-differentiable at $\mathbf{x}_0$, then the Hessian matrix at $\mathbf{x}_0$ is symmetric.

Theorem 5 (Young Theorem)**.** *Suppose $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$ is $\mathbf{C}^2$ at $\mathbf{x}_0 \in \operatorname{int}(X)$. Then,*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}_0)$$

whenever the cross partials exist.

Exercise 11 (PS8)**.** Prove Theorem 5 for the case when $d = 2$. **Hint**: Consider a rectangle formed with vertices at $\mathbf{x}_0$, $(x_{0,1} + h_1, x_{0,2})$, $(x_{0,1}, x_{0,2} + h_2)$, $(x_{0,1} + h_1, x_{0,2} + h_2)$. Let

$$\begin{aligned} r(\mathbf{h}) &\coloneqq f(x_{0,1} + h_1, x_{0,2} + h_2) - f(x_{0,1} + h_1, x_{0,2}), \\ t(\mathbf{h}) &\coloneqq f(x_{0,1} + h_1, x_{0,2} + h_2) - f(x_{0,1}, x_{0,2} + h_2) \end{aligned}$$

so that $r(\cdot)$ is the difference in f along the "right edge" of the rectangle and $t(\cdot)$ is the difference in f along the "top edge" of the rectangle. Let

$$d(\mathbf{h}) \coloneqq [f(x_{0,1} + h_1, x_{0,2} + h_2) - f(x_{0,1} + h_1, x_{0,2})] - [f(x_{0,1}, x_{0,2} + h_2) - f(\mathbf{x}_0)],$$

which is the difference in f along the right edge minus the difference along the left edge. Note that

$$d(\mathbf{h}) = r(h_1, h_2) - r(0, h_2) = t(h_1, h_2) - t(h_1, 0).$$

To proceed, apply the mean value theorem, re-express everything in terms of partials of f rather than partials of r and t, and then apply mean value theorem again. Divide both sides by $h_1 h_2$ to get almost what you want. Now take the limit of h_1 and h_2 to 0 and use continuity of the cross partials at $\mathbf{x}_0$ to conclude the result.

Exercise 12. Verify that below is an example of a function whose cross partials do not equal to one another. Why?

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

Definition 8. Let $\mathbf{v} \in \mathbb{R}^d$ be a unit-norm vector (i.e., $\|\mathbf{v}\| = 1$).[4] The directional derivative of

[4] We are interested in the direction of a vector only and thus its "length" measured by the norm is irrelevant. Put differently, we want to make sure that directional derivative in the direction $2\mathbf{v}$ and $\mathbf{v}$ to be the same at $\mathbf{x}$.

$f : X \subseteq \mathbb{R}^d \to \mathbb{R}^m$ in the direction $\mathbf{v}$ at $\mathbf{x}_0$ is defined as

$$D^{\mathbf{v}} f(\mathbf{x}_0) \coloneqq \begin{bmatrix} \lim_{h\to 0} \frac{f_1(\mathbf{x}_0 + h\mathbf{v}) - f_1(\mathbf{x}_0)}{h} \\ \vdots \\ \lim_{h\to 0} \frac{f_m(\mathbf{x}_0 + h\mathbf{v}) - f_m(\mathbf{x}_0)}{h} \end{bmatrix}.$$

Remark 14. Fix some $\mathbf{x}_0, \mathbf{v} \in \mathbb{R}^d$ with $\mathbf{v} \neq \mathbf{0}$ and define $g : \mathbb{R} \to \mathbb{R}^d$ as $g(t) \coloneqq \mathbf{x}_0 + t\mathbf{v}$. Then, $g(0) = \mathbf{x}_0$ and $g(t)$ gives a vector in the direction of $\mathbf{v}$. Suppose $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$ is differentiable on $\operatorname{int}(X)$. Define $h \coloneqq f \circ g$ and observe that

$$\begin{aligned} D^{\mathbf{v}} f(\mathbf{x}_0) &= \lim_{t\to 0} \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t} = \lim_{t\to 0} \frac{f(g(t)) - f(g(0))}{t} \\ &= \lim_{t\to 0} \frac{h(t) - h(0)}{t} = Dh(0) = D(f \circ g)(0). \end{aligned}$$

Then, the chain rule gives us that

$$D^{\mathbf{v}} f(\mathbf{x}_0) = (Df \circ g)(0)\, Dg(0) = Df(g(0))\, Dg(0) = \nabla f(\mathbf{x}_0)\, \mathbf{v}.$$

Since f is differentiable, by Proposition 11,

$$D^{\mathbf{v}} f(\mathbf{x}_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_d}(\mathbf{x}_0) \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_2 \end{bmatrix} = \sum_{i=1}^{d} \left(\frac{\partial f}{\partial x_i}(\mathbf{x}_0) \right) v_i.$$

Observe that, for any $i \in \{1, \ldots, d\}$,

$$D^{\mathbf{e}_i} f(\mathbf{x}_0) = \frac{\partial f}{\partial x_i}(\mathbf{x}_0).$$

Recall that the ith partial derivatives of f considers how the value of f changes when $\mathbf{x}_0$ moves in the direction of the ith coordinate. It follows that a directional derivative is a generalisation of the partial derivative when $\mathbf{x}_0$ is allowed to move in any arbitrary direction (specified by $\mathbf{v}$). You may now wonder if there is anything special about choosing $\mathbf{v} = \mathbf{e}_i$.

Proposition 13. *Suppose $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$ is differentiable at $\mathbf{x}_0 \in \operatorname{int}(X)$ and that $\nabla f(\mathbf{x}_0) \neq \mathbf{0}$. Then, the directional derivative $D^{\mathbf{v}} f(\mathbf{x}_0)$ is maximised when $\mathbf{v} = \nabla f(\mathbf{x}_0)/\|\nabla f(\mathbf{x}_0)\|$; i.e.,*

$$\frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|} = \operatorname*{arg\,max}_{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\| = 1} \|D^{\mathbf{v}} f(\mathbf{x}_0)\|$$

and the maximised directional derivative is $\|\nabla f(\mathbf{x}_0)\|$; i.e.,

$$\|\nabla f(\mathbf{x}_0)\| = \max_{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\| = 1} \|D^{\mathbf{v}} f(\mathbf{x}_0)\|.$$

Proof. We have

$$D^{\mathbf{v}} f(\mathbf{x}_0) = \nabla f(\mathbf{x}_0)\, \mathbf{v} = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}.$$

By Cauchy-Schwarz inequality, for any $\mathbf{v} \in \mathbb{R}^d$ with $\|\mathbf{v}\| = 1$, we have

$$|D^{\mathbf{v}} f(\mathbf{x}_0)| = |\nabla f(\mathbf{x}_0) \cdot \mathbf{v}| \leq \|\mathbf{v}\| \|\nabla f(\mathbf{x}_0)\| = \|\nabla f(\mathbf{x}_0)\|.$$

Since $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$, the inequality above holds with equality if and only if $\mathbf{v} = \nabla f(\mathbf{x}_0)/\|\nabla f(\mathbf{x}_0)\|$ because

$$\left|\nabla f(\mathbf{x}_0) \cdot \frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|}\right| = \frac{|\nabla f(\mathbf{x}_0) \cdot \nabla f(\mathbf{x}_0)|}{\|\nabla f(\mathbf{x}_0)\|} = \frac{\|\nabla f(\mathbf{x}_0)\|^2}{\|\nabla f(\mathbf{x}_0)\|} = \|\nabla f(\mathbf{x}_0)\|.$$

Finally, observe that

$$D^{\frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|}} f(\mathbf{x}_0) = \nabla f(\mathbf{x}_0) \cdot \frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|} = \|\nabla f(\mathbf{x}_0)\|. \qquad \blacksquare$$

Remark 15. If f is differentiable at $\mathbf{x}_0 \in \text{int}(X)$, then $\nabla f(\mathbf{x}_0)$ is equal to the vector of partial derivatives. Thus, proposition above tells us that directional derivative is largest (where size is measured by Euclidean norm $\|\cdot\|$) when moving in the direction of $\nabla f(\mathbf{x}_0)$.

4 Convexity

The following tells us something about the relationship between the derivative of a concave function at $\mathbf{x}$ and the slope between $\mathbf{x}$ and $\mathbf{y}$.

Proposition 14. *Let $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$ be concave and differentiable on X and suppose* $\text{int}(X)$ *is a convex set. Then, f is concave on* $\text{int}(X)$ *if and only if*

$$\nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geq f(\mathbf{y}) - f(\mathbf{x}) \;\; \forall \mathbf{x}, \mathbf{y} \in \text{int}(X).$$

Similarly, f is convex on $\text{int}(X)$ *if and only if*

$$\nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x}) \;\; \forall \mathbf{x}, \mathbf{y} \in \text{int}(X).$$

Proof. Suppose first that f is concave. Fix some $\mathbf{x}, \mathbf{y} \in \text{int}(X)$. Define $g : \mathbb{R} \to \mathbb{R}^d$ via $g(t) \coloneqq \mathbf{x} + t(\mathbf{y} - \mathbf{x})$. Since $g'(t) = \mathbf{y} - \mathbf{x}$, chain rule gives us that

$$(f \circ g)'(0) = \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}).$$

Moreover, by definition of the derivative,

$$(f \circ g)'(0) = \lim_{t \to 0} \frac{f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{t} = \lim_{t \searrow 0} \frac{f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{t}.$$

Since $\mathbf{x} + t(\mathbf{y} - \mathbf{x}) = t\mathbf{y} + (1-t)\mathbf{x}$, when $t \in (0,1)$, concavity of f of gives

$$\begin{aligned}\nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) &= \lim_{t \searrow 0} \frac{f(t\mathbf{y} + (1-t)\mathbf{x}) - f(\mathbf{x})}{t} \geq \lim_{t \searrow 0} \frac{tf(\mathbf{y}) + (1-t) f(\mathbf{x}) - f(\mathbf{x})}{t} \\ &= \lim_{t \searrow 0} (f(\mathbf{y}) - f(\mathbf{x})) = f(\mathbf{y}) - f(\mathbf{x}).\end{aligned}$$

To prove the converse, fix $\mathbf{x}, \mathbf{y} \in \operatorname{int}(X)$ and any $\alpha \in (0,1)$. Define

$$\mathbf{z} \coloneqq \alpha \mathbf{x} + (1-\alpha)\mathbf{y}, \;\; \mathbf{w} \coloneqq \mathbf{x} - \mathbf{z} = (1-\alpha)(\mathbf{x} - \mathbf{y}).$$

Note that $\mathbf{y} = \mathbf{z} - \frac{\alpha}{1-\alpha}\mathbf{w}$. By hypothesis,

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{z}) &\leq Df(\mathbf{z})(\mathbf{x}-\mathbf{z}) = Df(\mathbf{z})\mathbf{w}, \\ f(\mathbf{y}) - f(\mathbf{z}) &\leq Df(\mathbf{z})(\mathbf{y}-\mathbf{z}) = -\frac{\alpha}{1-\alpha}Df(\mathbf{z})\mathbf{w}. \end{aligned}$$

Multiplying the first inequality by $\frac{\alpha}{1-\alpha}$ and adding to the second gives

$$\begin{aligned} &\left[\frac{\alpha}{1-\alpha}f(\mathbf{x}) - \frac{\alpha}{1-\alpha}f(\mathbf{z})\right] + [f(\mathbf{y}) - f(\mathbf{z})] \leq \frac{\alpha}{1-\alpha}Df(\mathbf{z})\mathbf{w} - \frac{\alpha}{1-\alpha}Df(\mathbf{z})\mathbf{w} \\ &\Leftrightarrow \frac{\alpha}{1-\alpha}f(\mathbf{x}) + f(\mathbf{y}) - \frac{1}{1-\alpha}f(\mathbf{z}) \leq 0 \\ &\Leftrightarrow \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) \leq f(\mathbf{z}) = f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}); \end{aligned}$$

i.e., f is concave. Finally, the case for when f is convex follows by noting that $D(-f) = -Df$ and recalling that f is concave if and only if $-f$ is convex. ■

If a function is $\mathbf{C}^2$ (so that the Hessian f is symmetric), then concavity (or convexity) of functions can be characterised via the Hessian matrix (i.e., the second derivative). Recall that the Hessian of f at $\mathbf{x}$ is defined $H_f(\mathbf{x}) \coloneqq D^2 f(\mathbf{x})$.

Proposition 15. *Let $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$ be $\mathbf{C}^2$ on X and suppose $\operatorname{int}(X)$ is a convex set.*

(i) f is concave on $\operatorname{int}(X)$ if and only if $H_f(\mathbf{x})$ is negative semidefinite for all $\mathbf{x} \in \operatorname{int}(X)$.[5]

(ii) f is convex on $\operatorname{int}(X)$ if and only if $H_f(\mathbf{x})$ is positive semidefinite for all $\mathbf{x} \in \operatorname{int}(X)$.

(iii) If $H_f(\mathbf{x})$ is negative definite for all $\mathbf{x} \in \operatorname{int}(X)$, then f is strictly concave on $\operatorname{int}(X)$.

(iv) If $H_f(\mathbf{x})$ is positive definite for all $\mathbf{x} \in \operatorname{int}(X)$, then f is strictly convex on $\operatorname{int}(X)$.

Remark 16. Parts (iii) and (iv) implies that "strict" definiteness of the Hessian is only a sufficient condition for concavity/convexity.

We first prove Proposition 15 for the case when $d = 1$. Then, we prove a lemma that allows us to extend the $d = 1$ case to when $d > 1$.

Proposition 16. *Let $f : X \subseteq \mathbb{R} \to \mathbb{R}$ be $\mathbf{C}^2$ on $\operatorname{int}(X)$ and suppose $\operatorname{int}(X)$ is a convex set. Then, f is (resp. strictly) concave on $\operatorname{int}(X)$ if and only if $f''(x) \leq 0$ (resp. $f''(x) < 0$) for all $x \in \operatorname{int}(X)$.*

Proof. Let $x, y \in \operatorname{int}(X)$ with $x < y$. Pick sequences $(x_n)_n$ and $(y_n)_n$ in X such that $x < x_n < y_n < y$ for all $n \in \mathbb{N}$, $x_n \to x$ and $y_n \to y$. Concavity of f means that, for all $n \in \mathbb{N}$,

$$\frac{f(x) - f(x_n)}{x - x_n} \geq \frac{f(x_n) - f(y_n)}{x_n - y_n} \geq \frac{f(y_n) - f(y)}{y_n - y}. \tag{3}$$

[5] Recall that a symmetric matrix $M \in \mathbb{R}^{d\times d}$ is *negative* (resp. *positive*) *semidefinite* if $\mathbf{v}^\top M\mathbf{v} \leq 0$ (resp. $\mathbf{v}^\top M\mathbf{v} \geq 0$) for all $\mathbf{v} \in \mathbb{R}^d$. The matrix M is *negative* (resp. *positive*) *definite* if the inequality holds strictly for all $\mathbf{v} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$.

Observe that the left-most expression converges to $f'(x)$ while the right-most expression converges to $f'(y)$. Hence, $f'(x) \geq f'(y)$. Since x and y with $x < y$ were other chosen arbitrarily, it follows that f' is a nonincreasing function; i.e., its derivative, $f''(x)$, must be nonpositive for all $x \in \text{int}(X)$.

Conversely, suppose $f''(x) \leq 0$ for all $x \in \text{int}(X)$. Fix $x, y \in \text{int}(X)$ with $x < y$. Pick any $\alpha \in (0,1)$ and let $z = \alpha x + (1-\alpha)y$. By the Mean Value theorem, there exists $w_1 \in (x,z)$ and $w_2 \in (z,y)$ such that

$$\frac{f(x) - f(z)}{x - z} = f'(w_1) \; \frac{f(z) - f(y)}{z - y} = f'(w_2).$$

Since $w_1 < w_2$ and $f''(\cdot) \leq 0$, we must have $f'(w_1) \geq f'(w_2)$; i.e.,

$$\begin{aligned} \frac{f(x) - f(z)}{x - z} \geq \frac{f(z) - f(y)}{z - y} \Leftrightarrow f(z) &\geq \frac{y - z}{y - x} f(x) + \frac{z - x}{y - x} f(y) \\ &= \alpha f(x) + (1-\alpha) f(y); \end{aligned}$$

i.e., f is concave. The proof for strict concavity is analogous. ■

Exercise 13. Prove 3.

Exercise 14 (PS8)**.** Let $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$, where X is nonempty, open and convex. For any $\mathbf{x}, \mathbf{v} \in \mathbb{R}^d$, let $S_{\mathbf{x},\mathbf{v}} := \{t \in \mathbb{R} : \mathbf{x} + t\mathbf{v} \in X\}$ and define $g_{\mathbf{x},\mathbf{v}} : S_{\mathbf{x},\mathbf{v}} \to \mathbb{R}$ as $g_{\mathbf{x},\mathbf{v}}(t) := f(\mathbf{x} + t\mathbf{v})$. Then, f is (resp. strictly) concave on X if and only if $g_{\mathbf{x},\mathbf{v}}(\cdot)$ is (resp. strictly) concave for all $\mathbf{x}, \mathbf{v} \in \mathbb{R}^d$ with $\mathbf{v} \neq \mathbf{0}$.

Proof of Proposition **??**. For simplicity assume that X is open so that $X = \text{int}(X)$. We will prove part (i). Suppose f is concave and fix any $\mathbf{x}, \mathbf{v} \in X$. We wish to show that $\mathbf{v}^\top H_f(\mathbf{x})\mathbf{v} \leq 0$. Define $g_{\mathbf{x},\mathbf{v}}(t) = f(\mathbf{x} + t\mathbf{v})$ for all $t \in S_{\mathbf{x},\mathbf{v}}$ as in Exercise 14. Since f is $\mathbf{C}^2$, $g_{\mathbf{x},\mathbf{v}}$ is also $\mathbf{C}^2$ and so $g'_{\mathbf{x},\mathbf{v}}(t) = \nabla f(\mathbf{x} + t\mathbf{v})\mathbf{v}$ and $g''_{\mathbf{x},\mathbf{v}}(t) = \mathbf{v}^\top H_f(\mathbf{x} + t\mathbf{v})\mathbf{v}$. By Exercise 14, $g_{\mathbf{x},\mathbf{v}}(\cdot)$ is concave and so by Proposition 16, $g''_{\mathbf{x},\mathbf{v}}(t) \leq 0$ for all $t \in S_{\mathbf{x},\mathbf{v}}$. Therefore, $\mathbf{v}^\top H_f(\mathbf{x}+t\mathbf{v})\mathbf{v} \leq 0$ for all $t \in S_{\mathbf{x},\mathbf{v}}$. Because $0 \in S_{\mathbf{x},\mathbf{v}}$, we have $\mathbf{v}^\top H_f(\mathbf{x})\mathbf{v} \leq 0$.

Conversely, suppose that $H_f(\mathbf{x})$ is negative semidefinite for all $\mathbf{x} \in X$. By Exercise 14, to show hat f is concave, it suffices to show that, for any $\mathbf{x}, \mathbf{v} \in X$ with $\mathbf{v} \neq 0$, the function $g_{\mathbf{x},\mathbf{v}}(\cdot)$ is concave on $S_{\mathbf{x},\mathbf{v}}$. Since f is $\mathbf{C}^2$, we again have that $g'_{\mathbf{x},\mathbf{v}}(t) = \nabla f(\mathbf{x} + t\mathbf{v})\mathbf{v}$ and $g''_{\mathbf{x},\mathbf{v}}(t) = \mathbf{v}^\top H_f(\mathbf{x} + t\mathbf{v})\mathbf{v}$. Since $\mathbf{v}^\top H_f(\mathbf{z})\mathbf{v} \leq 0$ for all $\mathbf{z} \in X$ (why?), in particular, $\mathbf{v}^\top H_f(\mathbf{x} + t\mathbf{v})\mathbf{v} \leq 0$ for all $t \in S_{\mathbf{x},\mathbf{v}}$. Therefore, $g''_{\mathbf{x},\mathbf{v}}(\cdot) \leq 0$ and by Proposition 16, $g_{\mathbf{x},\mathbf{v}}$ is concave. ■

Exercise 15. Modify the proof for Proposition 15 (i) above to prove parts (ii), (iii) and (iv) of Proposition 15.

Exercise 16. Consider $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) := -x^4$. Show that f is strictly concave on $\mathbb{R}$ but that its Hessian is not positive definite.

Exercise 17 (PS8)**.** Let $f : \mathbb{R}^2_{++} \to \mathbb{R}$ be defined by $f(x,y) := x^\alpha y^\beta$ for some $\alpha, \beta > 0$. Compute the Hessian of f at $(x,y) \in \mathbb{R}^2_{++}$. Find conditions on α and β such that f is (i) strictly concave, (ii) f is concave but not strictly concave, (iii) f is neither concave nor convex.

Proof. Suppose first that $f : \mathbb{R}^2_{++} \to \mathbb{R}$. Note that

$$\nabla f(x,y) = \begin{bmatrix} \alpha x^{\alpha-1} y^{\beta} & \beta x^{\alpha} y^{\beta-1} \end{bmatrix}$$
$$H_f(x,y) = \begin{bmatrix} \alpha(\alpha-1) x^{\alpha-2} y^{\beta} & \alpha\beta x^{\alpha-1} y^{\beta-1} \\ \alpha\beta x^{\alpha-1} y^{\beta-1} & \beta(\beta-1) x^{\alpha} y^{\beta-2} \end{bmatrix} = \begin{bmatrix} -\alpha(1-\alpha)\frac{f(x,y)}{x^2} & \alpha\beta\frac{f(x,y)}{xy} \\ \alpha\beta\frac{f(x,y)}{xy} & -\beta(1-\beta)\frac{f(x,y)}{y^2} \end{bmatrix}.$$

Then,

$$\begin{aligned} \det(H_f(x,y)) &= \left[\alpha(\alpha-1)\frac{f(x,y)}{x^2}\beta(\beta-1)\frac{f(x,y)}{y^2}\right] - \left[\alpha\beta\frac{f(x,y)}{xy}\right]^2 \\ &= \alpha\beta\left[1-(\alpha+\beta)\right]\left(\frac{f(x,y)}{xy}\right)^2. \end{aligned}$$

Hence, the determinant is strictly positive if $\alpha+\beta<1$, zero if $\alpha+\beta=1$, and strictly negative if $\alpha+\beta>1$. Recall that a 2×2 matrix is $M=[x_{ij}]_{i,j\in\{1,2\}}$ is negative positive if and only if $x_{11}>0$ and $\det(X)>0$, and M is negative definite if and only if $x_{11}<0$ and $\det(X)>0$. Hence, if $\alpha+\beta\le 1$ (which implies $\alpha,\beta<1$), $H_f(x,y)$ is positive semidefinite so that f is concave. Moreover, if $\alpha+\beta<1$, $H_f(x,y)$ is in fact positive definite so that f is strictly concave. However, observe that if $\alpha+\beta>1$, then $H_f(x,y)$ is not positive/negative semi/definite. Hence, f is not concave nor convex. If the domain of f is $\mathbb{R}^2_+$, then $H_f(0,0)$ is not positive definite; however, f is still strictly concave. ■

Remark 17. The argument Exercise 13 gives the Cordal Slope lemma: if $f : X \subseteq \mathbb{R} \to \mathbb{R}$ is concave, $\mathrm{int}(X)$ is convex, then, for any $x_1, x_2, x_3 \in \mathrm{int}(X)$ such that $x_1 < x_2 < x_3$,[6]

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} \ge \frac{f(x_3)-f(x_1)}{x_3-x_1} \ge \frac{f(x_3)-f(x_2)}{x_3-x_2}.$$

You might see this again in the proof of Jensen's inequality.

5 Implicit Function Theorem

We have so far worked with functions $f : \mathbb{R}^d \to \mathbb{R}^k$. We can think of $\mathbf{x} \in \mathbb{R}^d$ as the input, or *exogenous variables*, and $\mathbf{y} = f(\mathbf{x})$ as the output, or *endogenous variables*. However, we cannot always separate these two types of variables and they could be be related through the function $g : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^k$:

$$g(\mathbf{x}, \mathbf{y}) = \mathbf{0}.$$

In such cases, we still want to ask how $\mathbf{y}$ varies as we change $\mathbf{x}$. In short, implicit function theorem tells us that, under some assumptions, if we can solve a system at a point, then we can solve the system in a neighbourhood of that point and that we can have expressions for the derivatives of the endogenous variables around that point.

Given a function $f : X \times Y \subseteq \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^k$, let the Jacobian of f with respect to $\mathbf{x} \in X$ at

[6] If f is strictly concave, then the inequalities hold strictly. If f is convex, then the inequalities in (**??**) are reversed.

$(\mathbf{x}_0, \mathbf{y}_0) \in \text{int}(X \times Y)$ be given by

$$D_{\mathbf{x}} f(\mathbf{x}_0, \mathbf{y}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial f_1}{\partial x_d}(\mathbf{x}_0, \mathbf{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial f_k}{\partial x_d}(\mathbf{x}_0, \mathbf{y}_0) \end{bmatrix}_{k \times d}.$$

Similarly, let the Jacobian of f with respect to $\mathbf{y} \in Y$ at $(\mathbf{x}_0, \mathbf{y}_0) \in \text{int}(X \times Y)$ be given by

$$D_{\mathbf{y}} f(\mathbf{x}_0, \mathbf{y}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial y_1}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial f_1}{\partial y_m}(\mathbf{x}_0, \mathbf{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial y_1}(\mathbf{x}_0, \mathbf{y}_0) & \cdots & \frac{\partial f_k}{\partial y_m}(\mathbf{x}_0, \mathbf{y}_0) \end{bmatrix}_{k \times m}.$$

In the following theorem, think of it as there being d unknowns, $k = m$ equations and m parameters.

Theorem 6 (Implicit function theorem)**.** *Suppose $f : X \times Y \subseteq \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^m$ is $\mathbf{C}^1$ and that $X \times Y$ is open. Let $(\mathbf{x}_0, \mathbf{y}_0) \in X \times Y$ be a point such that $f(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$ and $D_{\mathbf{y}} f(\mathbf{x}_0, \mathbf{y}_0)$ is invertible. Then, there exists an open ball $B_{\epsilon_X}(\mathbf{x}_0) \subseteq X$ around $\mathbf{x}_0$ and an open ball $B_{\epsilon_Y}(\mathbf{y}_0) \subseteq Y$ around $\mathbf{y}_0$ such that, for all $\mathbf{x} \in B_{\epsilon_X}(\mathbf{x}_0)$, there exists a unique $\mathbf{y} \in B_{\epsilon_Y}(\mathbf{y}_0)$ such that $f(\mathbf{x}, \mathbf{y}) = 0$. Therefore, the equation $f(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ implicitly defines a function $g : B_{\epsilon_X}(\mathbf{x}_0) \to B_{\epsilon_Y}(\mathbf{y}_0)$ with the property*

$$f(\mathbf{x}, g(\mathbf{x})) = \mathbf{0} \ \forall \mathbf{x} \in B_{\epsilon_X}(\mathbf{x}_0).$$

Moreover, g is differentiable at any $\mathbf{x} \in B_{\epsilon_X}(\mathbf{x}_0)$ and

$$Dg(\mathbf{x}) = -\left(D_{\mathbf{y}} f(\mathbf{x}, g(\mathbf{x}))\right)^{-1} D_{\mathbf{x}} f(x, g(x)).$$

Proof. The proof for the general case is beyond the scope of this class. However, let us prove the case for when $d = m = k = 1$. So let $(x_0, y_0) \in \text{int}(X \times Y)$ such that $f(x_0, y_0) = 0$. The hypothesis that $D_{\mathbf{y}} f(\mathbf{x}_0, \mathbf{y}_0)$ is invertible is equivalent to $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$. We will assume that $\frac{\partial f}{\partial y}(x_0, y_0) > 0$ while noting that the case in which $\frac{\partial f}{\partial y}(x_0, y_0) > 0$ can be proved analogously. We wish to show that there exists an open balls $B(x_0)$ and $B(y_0)$ such that, for all $x \in B(x_0)$, there is a unique $y \in B(y_0)$ such that $f(x, y) = 0$. We will then show that $g : B(x_0) \to B(y_0)$ defined by $f(\cdot, g(\cdot)) = 0$ is differentiable on $B(x_0)$ and that

$$g'(\cdot) = -\frac{\frac{\partial f(\cdot, g(\cdot))}{\partial x}}{\frac{\partial f(\cdot, g(\cdot))}{\partial y}}.$$

Lemma 1. *Suppose $h : Z \to \mathbb{R}$ is continuous and $h(z_0) > 0$ for some $z_0 \in \text{int}(Z)$. Then, there exists $\delta > 0$ such that $h(z) > 0$ for all $z \in B_\delta(z_0)$.*

Proof. Fix $\epsilon := \frac{f(z_0)}{2}$. By continuity, there exists $\delta > 0$ such that $|h(z) - h(z_0)| < \epsilon = \frac{f(z_0)}{2}$ for all $z \in B_\delta(z_0)$. If $h(z) < h(z_0)$, then this implies $h(z_0) - h(z) < \frac{h(z_0)}{2} \Leftrightarrow h(z) > \frac{h(z_0)}{2} > 0$. If $h(z) \geq h(z_0)$, then $h(z) > 0$ and so the result follows. ∎

Since f is $\mathbf{C}^1$, $\frac{\partial f}{\partial y}$ is continuous. By Lemma 6, there exists an open ball $B_\delta(x_0, y_0)$ for some $\delta > 0$ such that $\frac{\partial f}{\partial y}(x, y) > 0$ for all $(x, y) \in B_\delta(x_0, y_0)$. Let $\delta_X, \delta_Y > 0$ be such that $B_{\delta_X}(x_0) \times B_{\delta_Y}(y_0) \subseteq$

$B_\delta(x_0, y_0)$. Since $f(x_0, y_0) = 0$ and $f(x_0, \cdot)$ is strictly increasing on $B_{\delta_Y}(y_0)$, we have

$$f\left(x_0, y_0 + \frac{\delta_Y}{2}\right) > 0 > f\left(x_0, y_0 - \frac{\delta_Y}{2}\right).$$

Letting $\epsilon_Y := \frac{\delta_Y}{2}$, using Lemma 6 (twice—check this), we can find $\epsilon_X \in (0, \delta_X]$ so that $f(x, y_0 + \epsilon_Y) > 0 > f(x, y_0 - \epsilon_Y)$ for all $x \in B_{\epsilon_X}(x_0)$. For any $x \in B_{\epsilon_X}(x_0)$, since $f(x, y_0 - \epsilon_Y) > 0 > f(x, y_0 - \epsilon_Y)$, intermediate value theorem implies that there exists $y_x \in B_{\epsilon_Y}(y_0)$ such that $f(x, y_x) = 0$. Moreover, y_x is unique because $f(x, \cdot)$ is strictly increasing on $B_{\epsilon_Y}(y_0)$. Let $g : B_{\epsilon_X}(x_0) \to B_{\epsilon_Y}(y_0)$ be defined as $g(x) := y_x$.

We now prove that g is continuous at x_0. Observe that the argument above implies that

$$|y_x - y_0| = |g(x) - g(x_0)| < 2\epsilon_Y = \delta_Y \ \forall x \in B_{\epsilon_X}(x_0).$$

Fix $\epsilon > 0$ and suppose we had chosen $\epsilon_Y = \frac{\epsilon}{2}$ instead in the argument above. Then, we would have that for some $d \in (0, \epsilon_X]$, $|y_x - y_0| = |g(x) - g(x_0)| < \epsilon$ for all $x \in B_d(x_0)$. Hence, g is continuous at x_0. In fact, since for any $x \in B_{\epsilon_X}(x_0)$, $f(x, g(x)) = 0$ and $f(x, \cdot)$ is strictly increasing, we can repeat the same argument to conclude that g is continuous at any $x \in B_{\epsilon_X}(x_0)$.

To show that g is differentiable, note that, by construction, $f(x, g(x)) = 0$ for all $x \in B_{\epsilon_X}(x_0)$. Fix some $x \in B_{\epsilon_X}(x_0)$ and $\epsilon'_X = \epsilon_X - |x - x_0| > 0$. Since $x + \epsilon \in B_{\epsilon_X}(x_0)$,

$$\begin{aligned}
0 &= \underbrace{f(x + \epsilon'_X, g(x + \epsilon'_X))}_{=0} + \underbrace{f(x, g(x + \epsilon'_X)) - f(x, g(x + \epsilon'_X))}_{=0} - \underbrace{f(x, g(x))}_{=0} \\
&= [f(x + \epsilon'_X, g(x + \epsilon'_X)) - f(x, g(x + \epsilon'_X))] + [f(x, g(x + \epsilon'_X)) - f(x, g(x))].
\end{aligned}$$

Since f is differentiable, applying the mean value theorem to the terms in square brackets separately yields

$$\begin{aligned}
f(x + \epsilon'_X, g(x + \epsilon'_X)) - f(x, g(x + \epsilon'_X)) &= \frac{\partial f}{\partial x}(x_1, g(x + \epsilon'_X))\, \epsilon'_X, \\
f(x, g(x + \epsilon'_X)) - f(x, g(x)) &= \frac{\partial f}{\partial y}(x, y_1)\,(g(x + \epsilon'_X) - g(x))
\end{aligned}$$

for some $x_1 \in (x, x + \epsilon'_X)$ and y_1 between $g(x)$ and $g(x + \epsilon'_X)$. Thus, we have

$$\begin{aligned}
&0 = \frac{\partial f}{\partial x}(x_1, g(x + \epsilon'_X))\, \epsilon'_X + \frac{\partial f}{\partial y}(x, y_1)\,(g(x + \epsilon'_X) - g(x)) \\
\Leftrightarrow\ &\frac{g(x + \epsilon'_X) - g(x)}{\epsilon'_X} = -\frac{\frac{\partial f}{\partial x}(x_1, g(x + \epsilon'_X))}{\frac{\partial f}{\partial y}(x, y_1)}.
\end{aligned}$$

Taking limit as $\epsilon'_X \to 0$, we have $x_1 \to x$, $y_1 \to g(x)$ so that

$$g'(x) = \lim_{\epsilon'_x \to 0} \frac{g(x + \epsilon'_X) - g(x)}{\epsilon'_X} = -\frac{\frac{\partial f}{\partial x}(x, g(x))}{\frac{\partial f}{\partial y}(x, g(x))}. \qquad \blacksquare$$

Remark 18. Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable and that y is implicitly defined by x as follows:

$$f(x, y) = 0.$$

Let $(x_0, y_0) \in \mathbb{R}^2$ such that $f(x_0, y_0) = 0$. We want to ask how y_0 changes with x around x_0. Implicit function theorem tells us that we can find $g : (x_0 - \epsilon_X, x_0 + \epsilon_X) \to (y_0 - \epsilon_Y, y_0 + \epsilon_Y)$ for some $\epsilon_X, \epsilon_Y > 0$ such that

$$f\left(x, g\left(x\right)\right) = 0 \ \forall x \in \left(x_0 - \epsilon_X, x_0 + \epsilon_X\right).$$

Moreover, the theorem tells us that g is differentiable on $(x_0 - \epsilon_X, x_0 + \epsilon_X)$. If we define $h : (x_0 - \epsilon_X, x_0 + \epsilon_X) \to \mathbb{R}$ as

$$h\left(x\right) \coloneqq f\left(x, g\left(x\right)\right),$$

we can use the chain rule to to obtain

$$h'\left(x\right) = D_x f\left(x, g\left(x\right)\right) + D_y f\left(x, g\left(x\right)\right) g'\left(x\right),$$

By construction, $h(x) = 0$ for all $x \in (x_0 - \epsilon_X, x_0 + \epsilon_X)$ so that h is a constant function; i.e., $h'(x) = 0$ for all $x \in (x_0 - \epsilon_X, x_0 + \epsilon_X)$. Since $D_y f(x, g(x)) \neq 0$ by assumption, we have

$$g'\left(x\right) = -\frac{D_x f\left(x, g\left(x\right)\right)}{D_y f\left(x, g\left(x\right)\right)} = -\frac{\frac{\partial f}{\partial x}\left(x_0, y_0\right)}{\frac{\partial f}{\partial y}\left(x_0, y_0\right)} \ \forall x \in \left(x_0 - \epsilon_X, x_0 + \epsilon_X\right).$$

Remark 19. Given $f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, the set

$$C = \left\{(\mathbf{x}, y) \in \mathbb{R}^d \times \mathbb{R} : f\left(\mathbf{x}, y\right) = 0\right\}$$

defines a "curve" in $\mathbb{R}^d \times \mathbb{R}$. Suppose f is $\mathbf{C}^1$ and let $(\mathbf{x}_0, y_0) \in \mathbb{R}^d \times \mathbb{R}$ be a point on the curve; i.e., $f(\mathbf{x}_0, y_0) = 0$. If $\frac{\partial f}{\partial y}(\mathbf{x}_0, y_0) \neq 0$, the implicit function theorem gives us open sets $B_X \subseteq \mathbb{R}^d$ and $B_Y \subseteq \mathbb{R}$, and a function $g : B_X \subseteq \mathbb{R}^d \to B_Y \subseteq \mathbb{R}$ such that $g(\mathbf{x}_0) = y_0$ and $f(\mathbf{x}, g(\mathbf{x})) = 0$ for all $\mathbf{x} \in B_X$. Then, C contains the graph of the function g;

$$\operatorname{gr}\left(g\right) = \left\{(\mathbf{x}, y) \in B_X \times B_Y : g\left(\mathbf{x}\right) = y\right\} \subseteq C.$$

Moreover, $Dg(\mathbf{x}_0)$ gives the derivative of the tangent hyperplane to the curve at $(\mathbf{x}_0, y_0)$. The tangent hyperplane is given by

$$\begin{aligned} &\left\{(\mathbf{x}, y) \in \mathbb{R}^d \times \mathbb{R} : y - y_0 = Dg\left(\mathbf{x}_0\right)\left(\mathbf{x} - \mathbf{x}_0\right)\right\} \\ &= \left(-\mathbf{x}_0, -y_0\right) + \left\{(\mathbf{v}, \tilde{y}) \in \mathbb{R}^d \times \mathbb{R} : \tilde{y} = Dg\left(\mathbf{x}_0\right)\mathbf{v}\right\}, \end{aligned}$$

Because $Dg(\mathbf{x}_0)$ is a vector of partial derivatives, given any $\mathbf{v} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$, $Dg(\mathbf{x}_0)\mathbf{v}$ is the directional derivative of g at $\mathbf{x}_0$ in the direction of $\mathbf{v}$. Observe that

$$\begin{aligned} \nabla f\left(\mathbf{x}_0, y_0\right)\begin{bmatrix} \mathbf{v} \\ Dg\left(\mathbf{x}_0\right)\mathbf{v} \end{bmatrix} &= \begin{bmatrix} D_{\mathbf{x}} f\left(\mathbf{x}_0, y_0\right) & \frac{\partial f}{\partial y}\left(\mathbf{x}_0, y_0\right) \end{bmatrix}\begin{bmatrix} \mathbf{v} \\ -\frac{D_{\mathbf{x}} f(\mathbf{x}_0, y_0)}{\frac{\partial}{\partial y} f(\mathbf{x}_0, y_0)}\mathbf{v} \end{bmatrix} \\ &= D_{\mathbf{x}} f\left(\mathbf{x}_0, y_0\right)\mathbf{v} - \frac{\partial f}{\partial y}\left(\mathbf{x}_0, y_0\right)\frac{D_{\mathbf{x}} f\left(\mathbf{x}_0, y_0\right)}{\frac{\partial}{\partial y} f\left(\mathbf{x}_0, y_0\right)}\mathbf{v} \\ &= 0; \end{aligned}$$

i.e., the gradient vector is orthogonal to the tangent hyperplane to the curve at $(\mathbf{x}_0, y_0)$.

Exercise 18 (PS9)**.** Let $F : \mathbb{R}^4 \to \mathbb{R}^2$. Suppose the conditions for the implicit function theorem are satisfied at all points and that $F(x_1^*, x_2^*, y_1^*, y_2^*) = 0$. Let $h = (h_1, h_2)$ denote the implicitly defined function of (x_1, x_2) for the relation $F(x_1, x_2, y_1, y_2) = (0, 0)$ near $(x_1^*, x_2^*, y_1^*, y_2^*)$. Given explicit formulas for $\frac{\partial h_i}{\partial x_j}$ for $i, j \in \{1, 2\}$.

Corollary 4 (Inverse function theorem)**.** *Suppose $f : X \subseteq \mathbb{R}^d \to Y \subseteq \mathbb{R}^d$. Let $\mathbf{x}_0 \in \operatorname{int}(X)$ and define $\mathbf{y}_0 \coloneqq f(\mathbf{x}_0)$. If f is $\mathbf{C}^1$ and $Df(\mathbf{x}_0)$ is invertible, then there exists an open ball $B_{\epsilon_X}(\mathbf{x}_0) \subseteq X$ around $\mathbf{x}_0$ and an open ball $B_{\epsilon_Y}(\mathbf{y}_0) \subseteq Y$ around $\mathbf{y}_0$ such that, for all $\mathbf{y} \in B_{\epsilon_Y}(\mathbf{y}_0)$, there exists a unique $\mathbf{x} \in B_{\epsilon_X}(\mathbf{x}_0)$ such that $f(\mathbf{x}) = \mathbf{y}$. Therefore, the equation $f(\mathbf{x}) = \mathbf{y}$ implicitly defines a function $g : B_{\epsilon_Y}(\mathbf{y}_0) \to B_{\epsilon_X}(\mathbf{x}_0)$ with the property*

$$f\left(g\left(\mathbf{y}\right)\right) = \mathbf{y} \ \forall \mathbf{y} \in B_{\epsilon_Y}\left(\mathbf{y}_0\right).$$

Moreover, g is differentiable at any $\mathbf{y} \in B_{\epsilon_Y}(\mathbf{y}_0)$ and

$$Dg\left(\mathbf{y}\right) = \left(Df\left(g\left(\mathbf{y}\right)\right)\right)^{-1}.$$

Exercise 19 (PS9)**.** Prove Corollary 4. **Hint:** An inverse function of $f : X \to Y$, f^{-1}, satisfies following equation:

$$\mathbf{y} - f\left(f^{-1}\left(\mathbf{y}\right)\right) \equiv 0..$$

Thus, we can think of $\mathbf{x} = f^{-1}(\mathbf{z})$ as being implicitly defined (by $\mathbf{z}$) via the expression above.

Remark 20. Given $f : \mathbb{R} \to \mathbb{R}$ that is $\mathbf{C}^1$ with $f(x) = y$. The inverse function theorem tells us that for some $x_0 \in \mathbb{R}$ such that $f'(x_0) \neq 0$,

$$\left(f^{-1}\right)'\left(y_0\right) = \frac{1}{f'\left(x_0\right)} = \frac{1}{f'\left(f^{-1}\left(y_0\right)\right)},$$

where $y_0 \coloneqq f(x_0)$. Hence, we do not need the explicit expression of f^{-1} to compute the derivative of f^{-1}.