MATH20410 (W25): Analysis in Rn II (accelerated)

Lecturer: Joe Jackson
Notes by: Aden Chen

Wednesday 29th January, 2025

Contents

1 Single-Variable Differential Calculus

In this chapter, we consider mainly functions of the form $f : I \to \mathbb{R}$, where I is an interval, e.g., (a, b), $[a, b]$, $(a, b]$, (a, ∞), $\mathbb{R}$. This is the function we have in mind unless otherwise stated.

Definition 1.1 (Differentiability)**.** We say f is **differentiable at** $x \in I$ if the limit

$$f'(x) := \lim_{t \to x} \frac{f(t) - f(x)}{t - x} = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

exists. In this case, we call $f'(x)$ the derivative of f at x. Moreover:

- We say that f is **differentiable** if $f'(x)$ exists for each $x \in I$.

- We say f is **continuously differentiable** ($f \in C^1$) if $f' : I \to \mathbb{R}$ is continuous.

Example 1.2.

- $f(x) = |x|$. Differentiable on $\mathbb{R} \setminus \{0\}$.

- $f(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$. Continuous but not differentiable at 0.

- $f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$. Differentiable everywhere (in particular at 0), but $f \notin C^1$.

Proposition 1.3 (Rules for computing derivatives)**.**

(i) *Linearity.* $(af + bg)' = af' + bg'$ *(if f' and g' exist, such requirements are hereafter omitted).*

(ii) *Product rule.* $(fg)' = f'g + fg'$.

(iii) *Quotient rule.* $(f/g)' = (f'g - fg')/g^2$.[1]

(iv) *Chain rule.* $(f \circ g)' = (f' \circ g) \cdot g'$.

---

[1]Low dhigh minus high dlow. Not Haidilao...

Proof. We prove the quotient rule; the remaining are left as exercises. Starting from the definition

$$\left(\frac{f}{g}\right)'(x) = \lim_{t \to x} \frac{\frac{f}{g}(t) - \frac{f}{g}(x)}{t - x}$$

$$= \lim_{t \to x} \frac{\frac{f(t)}{f(t)} + \frac{f(x)}{g(t)} - \frac{f(x)}{g(t)} + \frac{f(x)}{g(x)}}{t - x}.$$

Note that

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

and we have

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g^2(x)}$$

□

Theorem 1.4. *If f is differentiable at x then f is continuous at x.*

Proof. Note that

$$\lim_{t \to x} f(t) - f(x) = \lim_{t \to x} \frac{f(t) - f(x)}{t - x}(t - x) = f'(x) \cdot 0 = 0.$$

□

1.1 The Mean Value Theorem

Lemma 1.5. *Suppose $f : [a, b] \to \mathbb{R}$ has a local maximum or minimum at $x \in (a, b)$. If $f'(x)$ exists, then $f'(x) = 0$.*

Proof. From the definition of the derivative, consider the limits from the left and right; one is non-positive and the other is non-negative. □

Theorem 1.6 (Rolle's Theorem)**.** *Suppose $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, differentiable on (a, b), and such that $f(a) = f(b)$. Then there exists $x \in (a, b)$ such that $f'(x) = 0$.*

Proof. Consider the global maximum or minimum (exist since f is a continuous function defined on a compact set) and apply the previous lemma. (If both the maximum and minimum is at a or b, f is constant.) □

4

Theorem 1.7 (Mean Value Theorem). *Let $f : [a, b] \to \mathbb{R}$ be such that f is continuous on $[a, b]$ and differentiable on (a, b). Then there exists $x \in (a, b)$ such that $f(b) - f(a) = f'(x)(b - a)$.*

Proof. Apply Rolle's to $\tilde{f} = f - [f(b) - f(a)] \cdot \frac{x-a}{b-a}$. □

Theorem 1.8. *Let $f : (a, b) \to \mathbb{R}$ be differentiable.*

(a) *if $f' = 0$, then f is constant.*

(b) *if $f' \geq 0$, then f is increasing.*

(c) *if $f' \leq 0$, then f is decreasing.*

Proof. Apply the mean value theorem. □

Theorem 1.9 (The Intermediate Value Property of Derivatives). *Let $f : [a, b] \to \mathbb{R}$ be differentiable[2] and suppose $f'(a) < \lambda < f'(b)$ Then there exists $x \in (a, b)$ such that $f'(x) = \lambda$.*

> [2] f need not be C^1!

Proof *(à la Pugh).* Slide a small secant of length so small that the slope around a and b is separated also by λ. By continuity of the slope, there exists a secant between a and b with slope λ. Apply the mean value theorem to this slope. □

Proof *(à la Joe/Rudin).* We start with $\lambda = 0$. Then $f'(a), f'(b) \neq 0$ and the global min/max of f cannot be at the endpoints. At the global extrema we have the desired result. When $\lambda \neq 0$, consider $\tilde{f} := f - \lambda x$. □

Example 1.10. Consider

$$f(x) = \begin{cases} x^2 \sin(1/x), & x \neq 0 \\ 0, & x = 0 \end{cases}.$$

We have

$$f(x) = \begin{cases} 2x \sin(1/x) = \cos(1/x), & x \neq 0 \\ 0, & x = 0 \end{cases},$$

which has the intermediate value property.

Theorem 1.11 (Generalized Mean Value Theorem). *Let $f, g : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $x \in (a, b)$ such that*

$$\big(f(a) - f(b)\big)g'(x) = \big(g(a) - g(b)\big)f'(x).$$

5

Remark 1.12. When the above is not zero,

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

Proof. Define

$$h(t) := \big(f(b) - f(a)\big)g(t) - \big(g(b) - g(a)\big)f(t).$$

Note that

$$h(a) = f(b)g(a) - f(a)g(b) = h(b)$$

and apply Rolle's. $\qquad\square$

1.2 L'Hôpital's Rule

Theorem 1.13 (L'Hôpital's Rule, a particular case)**.** *Let* $f, g : [a, b] \to \mathbb{R}$ *be continuous on* $[a, b]$ *and differentiable on* (a, b). *If* $g(x) \neq 0$ *in a neighborhood of* a *and* $f(x) = g(x) = 0$, *then*

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

if the last limit exists.

Proof. Consider some small $\delta > 0$. The generalized MVT gives some $x \in (a, a+\delta)$ such that

$$\frac{f(a + \delta)}{g(a + \delta)} = \frac{f'(x)}{g'(x)} \approx \lim_{t \to a} \frac{f'(t)}{g'(t)},$$

where the last approximation follows from the existence of the limit. Note that as $\delta \to 0$, $x \to a$, and the approximation error shrinks to 0. $\qquad\square$

Refer to Rudin or something for the general case.

1.3 Higher Derivatives

If $f : I \to \mathbb{R}$ is differentiable, then we can define the second derivative $f'' := (f')'$ if f' is differentiable. Higher derivatives can be defined similarly. We usually write $f^{(n)}$ for the n-th derivative of f.

Example 1.14. $L(x) = f(x_0) + f'(x_0)(x - x_0)$ is a (first order) linear approximation of f at x_0. How good is this approximation? A first answer is

$$f(x) = L(x) + o(|x - x_0|),$$

since we have as $x \to x_0$ that

$$\frac{f(x) - L(x)}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \longrightarrow 0.$$

But can we say even more about the quality of the approximation? – Yes, if f is twice differentiable.

Proposition 1.15 (First-order Taylor's Theorem). *Suppose f' exists and is continuous on $[a, b]$ and f'' exists on (a, b). Let $x_0, x \in [a, b]$ with $x_0 \neq x$. Then*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(y)(x - x_0)^2,$$

where y is between x_0 and x. In particular, we have

$$|f(x) - f(x_0) - f'(x_0)(x - x_0)| \leq \frac{1}{2} \sup_{y \in (a,b)} |f''(y)| \cdot |x - x_0|^2.$$

Proof. Find M such that we have

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{M}{2}(x - x_0)^2.$$

We need only find y such that $M = f''(y)$. Define

$$g(t) := f(t) - f(x_0) - f'(x_0)(t - x_0) - \frac{M}{2}(t - x_0)^2.$$

Note that $g''(t) = f''(t) - M$, so we need only find a point at which g'' vanishes. Since $g(x_0) = g(x) = 0$, by the MVT there exists y' between x_0 and x such that $g(y') = 0$. Observe that $g'(x_0) = 0$, and so by the MVT again, there exists y between x_0 and y' (and by extension between x_0 and x) such that $g''(y) = 0$. $\quad\square$

The more general story: given $f : [a, b] \to \mathbb{R}$ and $x_0 \in [a, b]$, we may define

$$P_0(x) := f(x_0),$$
$$P_1(x) := f(x_0) + f'(x_0)(x - x_0),$$
$$P_2(x) := f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2,$$
$$\vdots$$
$$P_n(x) := \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k,$$

when the corresponding derivatives exist. Note that $P_n(x)$ is the unique degree n polynomial such that $P_n^{(k)}(x_0) = f^{(k)}(x_0)$ for $k = 1, \ldots, n$.

Theorem 1.16 (Taylor's Theorem). *Let $f : [a, b] \to \mathbb{R}$ be such that*

- *$f^{(k)}$ exists on $[a, b]$ for $k = 1, \ldots, n$; and*

- *$f^{(n+1)}$ exists on (a, b).*

Then, for any $x_0, x \in [a, b]$ with $x_0 \neq x$, there exists y between x_0 and x such that

$$f(x) = P_n(x) + \frac{f^{(n+1)}(y)}{(n+1)!}(x - x_0)^{n+1}.$$

for some y between x_0 and x.

We proof the case $n = 2$, the same idea can be used to prove the general case.

Proof. Define
$$g(t) = f(t) - P_2(t) - \frac{M}{6}(t - x_0)^3.$$

Since $g''' = f''' - M$, we need only find y such that $g'''(y) = 0$. Note that $g(x_0) = g(x) = 0$, and so by the MVT there exists y' between x_0 and x such that $g'(y') = 0$. Next, note that $g'(x_0) = 0$, and so by the MVT there exists y'' between x_0 and y' such that $g''(y'') = 0$. Finally, note that $g''(x_0) = 0$, and so by the MVT there exists y between x_0 and y'' such that $g'''(y) = 0$. $\square$

2 Multivariable Differential Calculus

Some remainders about $\mathbb{R}^n$:

- $\mathbb{R}^n = \{x = (x_1, \ldots, x_n) : x_i \in \mathbb{R}\}$.

- $\mathbb{R}^n$ is a vector space, with canonical basis $\{e_i, \ldots, e_n\}$.

- $\mathbb{R}^n$ comes with an inner product $\langle x, y \rangle = x \cdot y = \sum x_i y_i$, a norm $|x| = \sqrt{x \cdot x} = (\sum x_i y_i)^{1/2}$, and a metric $d(x, y) = |x - y|$.

2.1 Higher Dimensional Codomains

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

Definition 2.1. f is differentiable at x if the limit

$$f'(x) := \lim_{t \to x} \frac{f(t) - f(x)}{t - x}$$

exists.

Remark 2.2. We may write $f(t) = (f_1(t), \ldots, f_n(t))$, and $f'(x) = (f_1'(x), \ldots, f_n'(x))$, since a sequence $x \in \mathbb{R}^n$ converges if and only if each of its components converges.

Theorem 2.3. *We have the following analog of the MVT:*

$$|f(b) - f(a)| \leq |f'(t)| \cdot |b - a|.$$

for some t between a and b.

Proof. Assume $a < b$. Define

$$h(t) := \langle f(b) - f(a), f(t) \rangle.$$

The MVT gives

$$h(b) - h(a) = h'(t)(b - a) = \langle f(b) - f(a), f'(t) \rangle (b - a)$$
$$\leq (b - a)|f(b) - f(a)||f'(t)|,$$

where the last inequality follows from the Cauchy-Schwarz inequality. Noting that

$$h(b) - h(a) = |f(b) - f(a)|^2,$$

we have the desired result. $\qquad \square$

2.2 Higher Dimensional Domain

We next consider functions $f : U \to \mathbb{R}$, where $U \subset \mathbb{R}^n$ is open.

Definition 2.4 (Partial Derivatives)**.**

$$\frac{\partial f}{\partial x_i}(x) = D_i f(x) := \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}.$$

Definition 2.5 (Directional Derivatives)**.** Fix $u \in \mathbb{R}^n$.

$$= D_i u f(x) := \lim_{h \to 0} \frac{f(x + hu) - f(x)}{h}.$$

2.2.1 The Derivative

Intuition: A function is differentiable if a first-order Taylor expansion holds. That is, if f is "well-approximated" by a linear function.

Definition 2.6. We denote the set of all linear maps from $\mathbb{R}^n$ to $\mathbb{R}$ as $L(\mathbb{R}^n, \mathbb{R})$.

Definition 2.7 (The Derivative)**.** A function f is differentiable at x if there exists a linear map $T \in L(\mathbb{R}^n, \mathbb{R})$ such that

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

In this case we write $Df(x) = T$. In other words, $f(x + h) = f(x) + Df(x)(h) + o(|h|)$.

Remark 2.8.

- If f is differentiable, then

$$Df : U \longrightarrow L(\mathbb{R}^n, \mathbb{R}).$$

- If is easy to check that Df is well defined, that is, there is at most one T such that the limit holds.

We may think of the linear map $T : \mathbb{R}^n \to \mathbb{R}$ as

$$T(u) = \langle u, v \rangle, \tag{1}$$

where $v := (Te_1, \ldots Te_n)$.

Definition 2.9 (The Gradient)**.** If f is differentiable at x, we define $\nabla f(x) = v$, where v satisfies (1). In other words,

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - \langle \nabla f(x), h \rangle}{|h|} = 0.$$

Theorem 2.10. *If f is differentiable at x, then $D_u f(x)$ exists for all $u \in \mathbb{R}^n$ and $D_u f(x) = Df(x)u = \langle \nabla f(x), u \rangle$.*

Proof. Note that as $t \to 0$, we have

$$\left| \frac{f(x+tu) - f(x)}{t} - Df(x)u \right| = \left| \frac{f(x+tu) - f(x) - Df(x)(tu)}{t} \right|$$

$$= \left| \frac{f(x+tu) - f(x) - Df(x)(tu)}{|tu|} \right| \cdot |u| \longrightarrow 0.$$

$\square$

Remark 2.11. In particular we have $D_i f(x) = D_{e_i} f(x) = Df(x)e_i = \langle \nabla f(x), e_i \rangle$. In other words, if f is differentiable, then $\nabla f(x) = (D_1 f, \ldots, D_n f)$.

Remark 2.12.

- Differentiability holds if and only if the gradient exists.

- Differentiability implies the existence of directional derivatives, which then implies the existence of partial derivatives. The converse implications are not true.

Example 2.13. Consider

$$f(x_1, x_2) := \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}.$$

It is easy to see that $D_1 f(0) = D_2 f(0) = 0$ but $D_{(1,1)} f(0)$ does not exist. Indeed, f is not even continuous on the line $t(1, 1)$.

Example 2.14. Consider

$$f(x_1, x_2) := \begin{cases} \frac{x_1^3}{x_1^2 + x_2^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}.$$

11

Note that
$$D_u f(0) = \lim_{t \to 0} \frac{t^3 u_1^3}{t^2 (u_1^2 + u_2^2)} \cdot \frac{1}{t} = \frac{u_1^3}{u_1^2 + u_2^2}.$$
However, $Df(0)$ cannot exist, since the above mapping is not linear.

Theorem 2.15. *If the partial derivatives $D_1 f, \ldots, D_n f$ exist and are continuous (in a neighborhood of x), then f is differentiable at x.*

Proof. Fix arbitrary $x \in E$ and define $Ah = \sum D_i f(x) h_i$. We write $\omega_k := \sum_{i=1}^{k} h_i e_i$ for $k = 1, \ldots, n$ and $\omega_0 := x$. Note that $\omega_n = h$. By the MVT we can find δ_k between 0 and h_k such that

$$f(x + h) - f(x) - Ah = \sum_{k=1}^{n} f(x + \omega_k) - f(x + \omega_{k-1}) - D_k f(x) h_k$$

$$= \sum_{k=1}^{n} h_k [D_k(x + \omega_k + \delta_i e_i) - D_k f(x)],$$

which by continuity of D_i is sublinear. $\qquad \square$

2.3   Extension to Functions with Higher Dimensional Codomains

Immediate.

We have

$$Df(x) \in L(\mathbb{R}^n, \mathbb{R}^m), \quad \mathbb{R}^n \ni h \longmapsto Df(x) \in L(\mathbb{R}^n, \mathbb{R}^m),$$

and
$$Df : \mathcal{U} \longmapsto L(\mathbb{R}^n, \mathbb{R}^m).$$

Note that we may identify $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ with a unique matrix $A = [Te_1, \ldots, Te_n]$ such that we have $Th = Ah$ for each h.

Definition 2.16. If f is differentiable at x, we can define $[Df(x)] \in \mathbb{R}^{n \times m}$ to be the unique matrix such that

$$Df(x)h = [Df(x)]h.$$

More generally, for $T \in L(\mathbb{R}^n, \mathbb{R}^m)$, we use $[T]$ to denote the corresponding matrix.

Theorem 2.17. *If $Df(x)$ exists, so do D_if_j, and we have*

$$[Df(x)] = [D_if_j] = [\nabla f_1(x) \dots \nabla f_m(x)]^\top.$$

It suffices to prove the following stronger proposition:

Proposition 2.18. *The function f is differentiable at x if and only if each f_i is differentiable at x. In this case,*

$$Df(x)h = (Df_1h, \dots, Df_m(x)h) = (\langle \nabla f_1(x), h \rangle, \dots, \langle \nabla f_m(x), h \rangle) = [Df(x)]h.$$

Proof. Suppose f_i is differentiable. Define $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ by the formula

$$Th = (Df_1h, \dots, Df_m(x)h).$$

Note that

$$\frac{|f(x+h) - f(x) - Th|}{|h|} = \left(\sum \frac{|f_i(x+h) - f_i(x) - Df_i(x)h|^2}{|h|} \right)^{1/2} \longrightarrow 0.$$

The other direction is left as an exercise. $\qquad\square$

Corollary 2.19. *If D_jf_i all exist and are continuous in a neighborhood of x, then f is differentiable at x.*

2.4   The Chain Rule

Consider

$$\mathbb{R}^n \supset \mathcal{U} \xrightarrow{g} \mathbb{R}^m \xrightarrow{f} \mathbb{R}^k.$$

Theorem 2.20 (Chain Rule). *If g is differentiable at x and f is differentiable at $g(x)$, then $f \circ g$ is differentiable at x and*

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x).$$

A formal calculation:[3] We have

$$
\begin{aligned}
f \circ g(x+h) &= f \circ g(x) + Df(g(x))\big(g(x+h) - g(x)\big) + o\big(g(x+h) - g(x)\big) \\
&= f \circ g(x) + Df(g(x))\big(Dg(x)h + o(|h|)\big) + o(|h|) \\
&= f \circ g(x) + Df(g(x))\big(Dg(x)h\big) + o(|h|).
\end{aligned}
$$

[3]In math, "formal calculation" often means calculation that is "systematic but without rigorous justification."

13

Proof. For small $h \in \mathbb{R}^p$, we write

$$g(x + h) = g(x) + Bh + R_g,$$

where $B = \mathrm{D}g(x)$ and $\lim_{h \to 0} R_g/h = 0$. Similarly, we write

$$f \circ g(x + h) = f(g(x) + Bh + R_g) = f \circ g(x) + ABh + AR_g + R_f,$$

where $A = \mathrm{D}f(g(x))$ and $\lim_{h \to 0} R_f/(Bh + R_g) \to 0$. It remains to note that the last two terms are sublinear. $\qquad\square$

2.5   Continuity of the Derivative

Let $f : \mathbb{R}^n \supset \mathcal{U} \to \mathbb{R}^M$, where $\mathcal{U}$ is open. Recall that if f is differentiable, we have defined

- $\mathcal{U} \ni x \to \mathrm{D}f(x) \in L(\mathbb{R}^n, \mathbb{R}^m)$.

- $\mathcal{U} \ni x \to [\mathrm{D}f(x)] \in \mathbb{R}^{m \times n}$.

- $\mathcal{U} \ni x \to \mathrm{D}_j f_i(x) \in \mathbb{R}, i = 1, \ldots, m, j = 1, \ldots, n$.

Definition 2.21. For $T \in (\mathbb{R}^n, \mathbb{R}^m)$, we define the operator norm

$$\|T\| = \sup_{|v|=1} |Tv| = \sup_{|v| \in \mathbb{R}^n \setminus \{0\}} \frac{|Tv|}{|v|}.$$

This gives rise to the standard norm induced metric: for $T, S \in L(\mathbb{R}^n, \mathbb{R}^m)$, we have

$$d(T, S) = \|T - S\|.$$

Definition 2.22. For $A \in \mathbb{R}^{m \times n}$, we define the operator norm $\|A\|_{\mathrm{op}} = \sup_{|v|} |Av|$. Thus $\|T\| = \|[A]\|_{\mathrm{op}}$.

Definition 2.23. For $A \in \mathbb{R}^{m \times n}$, we define the Frobenius norm $\|A\|_{\mathrm{F}} = \left(\sum_{i,j} A_{ij}^2 \right)^{1/2}$.

Proposition 2.24. *The following statements are equivalent:*

- $x \mapsto \mathrm{D}f(x)$ *is continuous (wrt d).*

- $x \mapsto [\mathrm{D}f(x)]$ *is continuous (wrt d_{op}).*

- $x \mapsto [\mathrm{D}f(x)]$ *is continuous (wrt d_F).*

- *Each $x \mapsto \mathrm{D}_j f_i(x)$ is continuous.*

Definition 2.25. The function f is C^1 if the above equivalent conditions hold.

2.6   The Inverse Function Theorem

Theorem 2.26 (The Inverse Function Theorem). *Let $f : \mathbb{R}^n \supset E \to \mathbb{R}^n$ be C^1, where E is open. Suppose $x_0 \in E$ and $\mathrm{D}f(x_0)$ is invertible. Then there exists a neighborhood U of x_0 such that f is a bijection from U to $V := f(U)$, and $f^{-1} : V \to U$ is C^1 with derivative $\mathrm{D}(f^{-1}(y)) = \left[\mathrm{D}f(f^{-1}(y))\right]^{-1}$.*

Remark 2.27.

- Thus if the first order Taylor expansion is invertible, then f is invertible locally.

- Consider the identities

$$x = f^{-1}(f(x)), \quad y = f(f^{-1}(y)).$$

  Differentiating

$$I = \mathrm{D}f^{-1}(f(x)) \circ \mathrm{D}f(x), \quad I = \mathrm{D}f(f^{-1}(y)) \circ \mathrm{D}f^{-1}(y).$$

  This shows that $\mathrm{D}(f^{-1}(y))$ and $\mathrm{D}f(f^{-1}(x))$ are inverses of each other, provided that the functions are differentiable.

- Remember the one-dimensional case! We have that $(f^{-1})' = 1/f'$:

Proof *(Inverse Function Theorem, $n = 1$).* Let $\mathrm{D}f(x_0) \in L(\mathbb{R}, \mathbb{R})$ be invertible. Then $f'(x_0) \neq 0$, say $f'(x_0) > 0$ without loss of generality. By continuity of f', there exists an open interval U containing x_0 such that $f' > 0$ on U. Thus f is strictly increasing and thus one-to-one on U. It is easy to verify that $V := f(U) = (f(a), f(b))$, so V is open.

Next, we show that f^{-1} is continuous. For that, consider sequence $y_k \to y$. We seek to show that $f^{-1}(y_k) \to f^{-1}(y)$. Equivalently, given $f(x_k) \to f(x)$, we show $x_k \to x$. To that end, suppose not. Then, without loss of generality, there exists infinitely many x_k such that $x_k > x + \epsilon$ for some ϵ. Thus $f(x_k) > f(x + \epsilon) > f(x)$, a contradiction.

Finally, we show that f^{-1} is differentiable. Write $x := f^{-1}(y)$ and $f^{-1}(y + h) = x + k$, that is, define $k := f^{-1}(y+h) - f^{-1}(y)$. We have then that $h = f(x+k) - f(x)$. Then as $h \to 0$, we have $\lim_{h \to 0} k = 0$, by the continuity of f^{-1}, and so

$$\frac{f^{-1}(y+h) - f^{-1}(y)}{h} = \frac{k}{f(x+h) - f(x)} \longrightarrow \frac{1}{f'(x)}.$$

$\square$

Before the general proof, we need the following result:

Theorem 2.28 (Contraction Mapping)**.** *Let (X, d) be a complete metric space. Let* $\phi : X \to X$ *be a* ***contraction***, *that is, there exists $c < 1$ such that*

$$d\big(\phi(x), \phi(y)\big) \leq cd(x, y).$$

Then, there is a unique fixed point of ϕ.

Proof. Pick any $x_0 \in X$. Define $x_n := \phi(x_{n-1})$ for $n \geq 1$. Note that

$$\phi\big(x_n, x_{n-1}\big) \leq c^n \phi\big(x_1, x_0\big).$$

Thus, for $n > m$, we have

$$d\big(x_n, x_m\big) \leq \sum_{k=m+1}^{n} d\big(x_k, x_{k-1}\big) \leq d\big(x_1, x_0\big) \sum_{k=m+1}^{n} c^{k-1}.$$

Since $\sum c^j$ is a converging series, the last term tends to 0 and so (x_n) is Cauchy. Then, setting $x = \lim x_n$, we have

$$\phi(x) = \lim \phi(x_n) = \lim x_{n+1} = x.$$

Uniqueness follows from the contraction property. $\qquad\qquad\square$

We may now proceed with the general proof of the Inverse Function Theorem. We recall first the result:

Theorem 2.29 (The Inverse Function Theorem)**.** *Let $f : \mathbb{R}^n \supset E \to \mathbb{R}^n$ be C^1, where E is open. Suppose $x_0 \in E$ and $\mathrm{D}f(x_0)$ is invertible. Then there exists a neighborhood U of x_0 such that f is a bijection from U to $V := f(U)$, and $f^{-1} : V \to U$ is C^1 with derivative $\mathrm{D}(f^{-1}(y)) = \big[\mathrm{D}f(f^{-1}(y))\big]^{-1}$.*

Proof *(Inverse Function Theorem, the General Case)***.**
    Step 1: Local Invertibility. Choose δ small enough that

- $\big\|\mathrm{D}f(x)^{-1}\big\|$ is bounded in $B_\delta(x_0)$.[4]

- $\|\mathrm{D}f(x) - \mathrm{D}f(x')\|$ is "really small" if $x, x' \in B_\delta(x_0)$.

[4]Here, we used the fact that inversion is a continuous operation.

16

We check that f is injective on $U := B_\delta(x)$. Note that $f(x) = y$ if and only if $Df(x_0)^{-1}(y - f(x)) = 0$, which is equivalent to x being a fixed point of the function

$$\phi_y(x) := x + Df(x_0)^{-1}(y - f(x)).$$

Thus, to prove injectivity, we need only show that ϕ_y is a contraction. Observe that

$$D\phi_y(x) = I - Df(x_0)^{-1}Df(x) = Df(x_0)^{-1}\left[Df(x_0) - Df(x)\right].$$

Then,

$$\left\|D\phi_y(x)\right\| \le \left\|Df(x_0)^{-1}\right\|\left\|Df(x_0) - Df(x)\right\|$$

can be made arbitrarily small, and in particular smaller than $1/2$, by choosing δ small enough. The function ϕ_y is then a contraction. While the image of ϕ_y may not be a subset of its domain U (and so Banach contraction does not apply), the same argument in the proof of the Banach contraction theorem shows that ϕ_y has at most one fixed point, if any, in U. Injectivity of f in U thus follows.

Set $V := f(U)$. Note that f^{-1} is well defined on V.

Step 2: V is open. Fix $f(x_0) \in V$. Pick $r > 0$ such that $B_r(x_0) \subset U$. Note that

$$|x - x_0| \le \left\|Df(x_0)^{-1}\right\||f(x) - f(x_0)|.$$

Thus for $y = f(x)$ within $r/2\left\|Df(x_0)^{-1}\right\|$ of $f(x_0)$, we have $x \in U$ and so $y \in V$.

Step 3: f^{-1} is continuous (Lipschitz). Recall that $\phi_y(x)$ is a contraction in x with Lipschitz constant $1/2$, and note that it is also Lipschitz in y, with Lipschitz constant say C. From

$$x - x' = \phi_y(x) - \phi_{y'}(x') = \phi_y(x) - \phi_y(x') + \phi_y(x') - \phi_{y'}(x')$$

we thus know

$$|x - x'| \le \frac{1}{2}|x - x'| + C|y - y'|.$$

Then,

$$\left|f^{-1}(y) - f^{-1}(y')\right| = |x - x'| \le 2C|y - y'|$$

and f^{-1} is Lipschitz.

Step 4: The formula for $\mathrm{D}f^{-1}$**.** Write $y = f(x)$. Set $h = f^{-1}(y+k) - f^{-1}(y)$. Note that $f^{-1}(y+k) = x+h$ and so $k = f(x+h) - f(x)$. We have then that

$$
\frac{\left| f^{-1}(y+k) - f^{-1}(y) - \mathrm{D}f(x)^{-1}k \right|}{|k|}
$$
$$
= \frac{\left| h - \mathrm{D}f(x)^{-1}\big(f(x+h) - f(x)\big) \right|}{|f(x+h) - f(x)|}
$$
$$
\leq \frac{\left\| \mathrm{D}f(x)^{-1} \right\| \left\| \mathrm{D}f(x)h - f(x+h) + f(x) \right\|}{|h|} \cdot \frac{|h|}{|f(x+h) - f(x)|}.
$$

Note that the first term tends to 0 and the second is bounded. We have established then that that $\mathrm{D}f^{-1}(y) = \mathrm{D}f(x)^{-1}$ is continuous. It remains to note that as a composition of continuous functions, $\mathrm{D}f^{-1}$ is continuous. $\qquad\square$