MAT 322 Supplementary Notes

1 Examples of Topological Spaces

Let X be a set. A **topology** on X is a subset $\mathcal{T}$ of the power set $\mathcal{P}(X) = 2^X$ satisfying all of the following properties.

(i) The empty set is an element of $\mathcal{T}$, and X is an element of $\mathcal{T}$.

(ii) For every pair $(U, V) \in \mathcal{T} \cap \mathcal{T}$, also the intersection $U \cap V$ is in $\mathcal{T}$.

(iii) For every subset $\{U_\alpha\}_{\alpha \in I}$ of $\mathcal{T}$, also the union $\cup_{\alpha \in I} U_\alpha$ is in $\mathcal{T}$.

The subsets of X that are elements of $\mathcal{T}$ are called **$\mathcal{T}$-open** subsets of X. Thus (i) says that $\emptyset$ and X are $\mathcal{T}$-open subsets of X. Also (ii) says that the intersection of any two $\mathcal{T}$-open subsets of X is again a $\mathcal{T}$-open subset of X. Finally, (iii) says that the union of an arbitrary collection of $\mathcal{T}$-open subsets of X is a $\mathcal{T}$-open subset of X. A **topological space** is a pair $(X, \mathcal{T})$ of a set X and a topology $\mathcal{T}$ on X.

For topological spaces $(X, \mathcal{T})$ and $(Y, \mathcal{S})$, a function $f : X \to Y$ is **continuous** (with respect to $\mathcal{T}$ and $\mathcal{S}$) if for every $\mathcal{S}$-open subset $V \subset Y$, the inverse image subset $f^{-1}(V) \subset X$ is $\mathcal{T}$-open. To avoid the words "with respect to", often the topologies are denoted as follows.

$$f : (X, \mathcal{T}) \to (Y, \mathcal{S}).$$

In particular, for topologies $\mathcal{T}$ and $\mathcal{S}$ on a common set X, if the identity function

$$\mathrm{Id}_X : (X, \mathcal{T}) \to (X, \mathcal{S})$$

is continuous, then $\mathcal{T}$ **refines** $\mathcal{S}$, the topology $\mathcal{T}$ is **finer** than (or, better, "at least as fine as") the topology $\mathcal{S}$, and $\mathcal{S}$ is **coarser** than (or, better, "at least as coarse as") the topology $\mathcal{T}$. Although it sounds unusual in English, this does mean that every topology refines itself.

The study of topological spaces is a huge subject with connections to every other branch of mathematics. The examples below are just a sampling (quite idiosyncratic, but relevant for our course).

Example 1. Discrete Topology. The **discrete topology** on X is all of $\mathcal{P}(X)$, i.e., every subset of X is an open subset for the discrete topology. The discrete topology refines every topology on X. For every topological space $(Y, \mathcal{S})$, for every function $f : X \to Y$, the function

$$f : (X, \mathcal{P}(X)) \to (Y, \mathcal{S})$$

is continuous.

Example 2. Indiscrete Topology. The **indiscrete topology** on X is $\mathrm{I}_X = \{\emptyset, X\}$. Every topology on X refines the indiscrete topology. For every topological space $(Y, \mathcal{S})$, for every function $f : Y \to X$, the function

$$f : (Y, \mathcal{S}) \to (X, \mathrm{I}_X)$$

is continuous.

Example 3. Finite Complement Topology. The **finite complement** topology on X, $\mathrm{FC}_X \subset \mathcal{P}(X)$ consists of $\emptyset$ together will all subsets $U \subset X$ such that the complementary subset $X \setminus A$ is a finite set. If X is a finite set, this topology agrees with the discrete topology. For every function $f : X \to Y$ whose fiber sets are finite,

$$f : (X, \mathrm{FC}_X) \to (Y, \mathrm{FC}_Y)$$

is continuous.

Example 4. Metric Topology. For every metric space $(X, d : X \times X \to \mathbb{R})$, the d**-metric topology**, $\mathcal{T}_d \subset \mathcal{P}(X)$, consists of all subsets $U \subset X$ such that for every x in U, there exists real $\delta > 0$ (depending on U and x) such that the open ball of radius δ centered at x, $U^d_\delta(x)$, is contained in U. By Theorem 3.1, this is a topology on X. For metric spaces (X, d) and (Y, e), a function

$$f : (X, \mathcal{T}_d) \to (Y, \mathcal{T}_e)$$

is continuous if and only if it satisfies the ϵ-δ definition, i.e., for every $x \in X$, for every real $\epsilon > 0$, there exists real $\delta > 0$ such that $U^d_\delta(x)$ is contained in $f^{-1}(U^e_\epsilon(f(x)))$.

Example 5. Inverse Image Topology. Let $(X, \mathcal{T})$ be a topological space. Let A be a set, and let $i : A \to X$ be a function. The **inverse image topology** on A, $i^{-1}\mathcal{T}\mathcal{P}(A)$ consists of all subsets of the form $i^{-1}(U)$ where $U \subset X$ is in $\mathcal{T}$. In particular,

$$i : (A, i^{-1}(\mathcal{T})) \to (X, \mathcal{T})$$

is continuous. For every topological space $(Y, \mathcal{S})$, for every function $f : Y \to A$,

$$f : (Y, \mathcal{S}) \to (A, i^{-1}(\mathcal{T}))$$

is continuous if and only if the composite

$$i \circ f : (Y, \mathcal{S}) \to (X, \mathcal{T})$$

is continuous. In particular, for the inclusion function $i : A \to X$ of a subset A of X, the inverse image topology is called the **subspace topology** on A induced by $\mathcal{T}$, denoted $\mathcal{T}|A$. For for every topology $\mathcal{R}$ on A such that

$$i : (A, \mathcal{R}) \to (X, \mathcal{T})$$

is continuous, $\mathcal{R}$ refines $\mathcal{T}|A$.

Example 6. Product Topology. This is the definition for a product of two factors, although this extends easily to the setting of a product of infinitely many factors. Also this example can be combined with the previous example to define "(inverse) limit topologies". Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces. For the Cartesian product set $X \times Y$, a subset $W \subset X \times Y$ is $\mathcal{T} \times \mathcal{S}$-open if for every $(x, y) \in W$, there exists a $\mathcal{T}$-open subset $U \subset X$ containing x and a $\mathcal{S}$-open subset $V \subset Y$ containing y such that $U \times V$ is contained in W. Stated differently, the **product topology**, $\mathcal{T} \widehat{\times} \mathcal{S} \subset \mathcal{P}(X \times Y)$, is the collection of arbitrary unions of sets of the form $U \times V$ for $U \in \mathcal{T}$ and $V \in \mathcal{S}$. The projection function $\text{pr}_1 : X \times Y \to X$, respectively $\text{pr}_2 : X \times Y \to Y$, gives a continuous function,

$$\text{pr}_1 : (X \times Y, \mathcal{T} \widehat{\times} \mathcal{S}) \to (X, \mathcal{T}),$$

respectively,

$$\text{pr}_2 : (X \times Y, \mathcal{T} \widehat{\times} \mathcal{S}) \to (Y, \mathcal{S}).$$

For every topological space $(Z, \mathcal{R})$, for every function $f : Z \to X \times Y$,

$$f : (Z, \mathcal{R}) \to (X \times Y, \mathcal{T} \widehat{\times} \mathcal{S})$$

is continuous if and only if *both* the composite

$$\text{pr}_1 \circ f : (Z, \mathcal{R}) \to (X, \mathcal{T}),$$

and the composite

$$\text{pr}_2 \circ f : (Z, \mathcal{R}) \to (Y, \mathcal{S}),$$

are continuous.

Example 7. Pushforward Topology. Let $(X, \mathcal{T})$ be a topological space. Let A be a set, and let $q : X \to A$ be a function. The **pushforward topology** on A, $q_*\mathcal{T}$, consists of all subsets $V \subset A$ such that $q^{-1}(V) \subset X$ is in $\mathcal{T}$. In particular,

$$q : (X, \mathcal{T}) \to (A, q_*\mathcal{T})$$

is continuous. For every topological space $(Y, \mathcal{S})$, for every function $f : A \to Y$,

$$f : (A, q_*\mathcal{T}) \to (Y, \mathcal{S})$$

is continuous if and only if the composite

$$f \circ q : (X, \mathcal{T}) \to (Y, \mathcal{S})$$

is continuous. In particular, for the quotient map $q : X \to A$ corresponding to a partition / equivalence relation on X, the pushforward topology is called the **quotient topology** on A. For every topology $\mathcal{R}$ on A such that

$$q : (X, \mathcal{T}) \to (A, \mathcal{R})$$

is continuous, $q_*\mathcal{T}$ refines R.

2 Comparison of Metrics and Metric Topologies.

Let X be a set. Let d and e be metric functions on X, i.e., $X \times X \to \mathbb{R}$ satisfying the axioms for a metric space. Let C be a positive real number. The metric e is **C-Lipschitz** with respect to d, written $e \preceq_C d$, if for every $(x, y) \in X \times X$, $e(x, y) \leq C \cdot d(x, y)$. If there exists a positive real number C such that e is C-Lipschitz with respect to d, then e is **Lipschitz** with respect to d, $e \preceq d$. (Technically, it is the identity function $\mathrm{Id}_X : (X, d) \to (X, e)$ that is Lipschitz, but the notion above is relevant here.) The relation $\preceq$ on metric functions is transitive, but it is *not* a partial order because it is not antisymmetric. Instead, it induces an equivalence relation on metric functions. The metrics d and e are **equivalent** if both $d \preceq e$ and $e \preceq d$. Explicitly, d and e are equivalent if there exist real numbers $B > 0$ and $C > 0$ such that for every $(x, y) \in X \times X$, both $e(x, y) \leq C \cdot d(x, y)$ and $d(x, y) \leq B \cdot e(x, y)$. The property of being Lipschitz induces a partial order on the equivalence classes of equivalent metrics.

Proposition 2.1. *If e is Lipschitz with respect to d, then $Id_X : (X, d) \to (X, e)$ is continuous, and even uniformly continuous. In particular, if d and e are equivalent metrics, then the metric topology $\mathcal{T}_e$ equals the metric topology $\mathcal{T}_d$.*

Proof. For every real $\epsilon > 0$, set $\delta = \epsilon/C$, which is again a positive real. If $d(x, y) < \delta$, then $e(x, y) < \epsilon$. Thus Id_X is uniformly continuous. In particular, every $\mathcal{T}_e$ is contained in $\mathcal{T}_d$. If d and e are equivalent, i.e., if also d is Lipschitz with respect to e, then also $\mathcal{T}_d$ is contained in $\mathcal{T}_e$. Thus $\mathcal{T}_d$ equals $\mathcal{T}_e$. ☐

Let $(X, d : X \times X \to \mathbb{R})$ be a metric space. For every subset $A \subset X$, the **restriction metric** on A, $d|A$, is simply the restriction of d to $A \times A$. By Theorem 3.2, the metric topology on A induced by the restriction metric $d|A$ equals the subspace topology on A induced by the metric topology on X.

Let (X, d) and (Y, e) be metric spaces. There are several natural definitions of an associated metric on $X \times Y$, but they are all equivalent. The **supremum product metric** on $X \times Y$ is

$$\rho : (X \times Y) \times (X \times Y) \to \mathbb{R}, \ \rho((x, y), (x', y')) := \sup(d(x, x'), e(y, y')).$$

In particular, $U^{\rho}_{\epsilon}((x, y))$ equals $U^{d}_{\epsilon}(x) \times U^{e}_{\epsilon}(y)$.

Proposition 2.2. *The metric topology $\mathcal{T}_\rho$ on $X \times Y$ equals the product topology $\mathcal{T}_d \widehat{\times} \mathcal{T}_e$.*

Proof. Let $W \subset X \times Y$ be a $\mathcal{T}_\rho$-open subset. For every element $(x,y) \in W$, there exists real $\epsilon > 0$ such that $U^\rho_\epsilon((x,y))$ is contained in W. Now $U^\rho_\epsilon((x,y))$ equals $U^d_\epsilon(x) \times U^e_\epsilon(y)$. Since $U^d_\epsilon(x)$ is a $\mathcal{T}_d$-open neighborhood of x in X, and Since $U^e_\epsilon(y)$ is a $\mathcal{T}_e$-open neighborhood of y in Y, it follows that W is open in the product topology. Thus every $\mathcal{T}_\rho$-open subset of $X \times Y$ is open in the product topology.

Conversely, for every subset $W \subset X \times Y$ that is open in the product topology, for every (x,y) in W, there exists a $\mathcal{T}_d$-open neighborhood U of x in X and there exists a $\mathcal{T}_e$-open neighborhood V of y in Y such that $U \times V$ is contained in W. Since U is $\mathcal{T}_d$-open, there exists real $\delta > 0$ such that $U^d_\delta(x)$ is contained in U. Since V is $\mathcal{T}_e$-open, there exists real $\epsilon > 0$ such that $U^e_\epsilon(y)$ is contained in V. Thus, for the real number $\lambda = \min(\delta, \epsilon)$, λ is positive and $U^\rho_\lambda((x,y))$, which equals $U^d_\lambda(x) \times U^e_\lambda(y)$, is contained in $U^d_\delta(x) \times U^e_\epsilon(y)$, which in turn is contained in $U \times V$, which in turn is contained in W. Thus W is $\mathcal{T}_\rho$-open. Thus every subset of $X \times Y$ that is open in the product topology is $\mathcal{T}_\rho$-open. □

The two other common product metrics are the ℓ_1-product metric,

$$\rho_1((x,y),(x',y')) = d(x,x') + e(y,y'),$$

and the Euclidean product metric,

$$\rho_E((x,y),(x',y')) = \sqrt{d(x,x')^2 + e(y,y')^2}.$$

These are all equivalent,

$$\rho((x,y),(x',y')) \leq \rho_1((x,y),(x',y')) \leq 2\rho((x,y),(x',y')),$$
$$\rho((x,y),(x',y')) \leq \rho_E((x,y),(x',y')) \leq \sqrt{2}\rho((x,y),(x',y')).$$

Let $(V,+,0,\cdot)$ be a $\mathbb{R}$-vector space. Let $\| \bullet \|$ and $| \bullet |$ be norm functions on V, i.e., functions $V \to \mathbb{R}$ satisfying the axioms for a norm. Let C be a positive real number. The norm $\| \bullet \|$ is **C-Lipschitz**, resp. **Lipschitz**, with respect to $| \bullet |$ if the corresponding metric function $e = d_{\|\bullet\|}$ is C-Lipschitz, resp. Lipschitz, with respect to the metric function $d = d_{|\bullet|}$. Because of the axioms of norm functions, this is equivalent to the following: for every $\mathbf{x} \in V$, $\|\mathbf{x}\| \leq C \cdot |\mathbf{x}|$. If both $\| \bullet \|$ is Lipschitz with respect to $| \bullet |$, and $| \bullet |$ is Lipschitz with respect to $\| \bullet \|$, then $\| \bullet \|$ is **equivalent** to $| \bullet |$.

For $\mathbb{R}^d$, there are many inequivalent metric functions. However, all norms on this vector space are equivalent. Recall that the ℓ_1-norm on $\mathbb{R}^d$ is defined by,

$$\|\vec{v}\| := |x_1| + \cdots + |x_n|, \ \vec{v} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Proposition 2.3. *Every norm $\| \bullet \|$ on $\mathbb{R}^d$ is equivalent to the ℓ_1-norm $\| \bullet \|_1$.*

Proof. Denote by C the maximum of $\| \bullet \|$ on the standard basis set for $\mathbb{R}^d$,

$$C = \max(\|\mathbf{e}_1\|, \dots, \|\mathbf{e}_n\|).$$

Since $\| \bullet \|$ is positive definite, C is a positive real number. By the triangle inequality and homogeneity of $\| \bullet \|$,

$$\|\vec{v}\| \leq |x_1| \cdot \|\mathbf{e}_1\| + \cdots + |x_n| \cdot \|\mathbf{e}_n\| \leq C \cdot \|\vec{v}\|_1.$$

Since $\| \bullet \|$ is Lipschitz with respect to $\| \bullet \|_1$, the induced function,

$$\| \bullet \| : (\mathbb{R}^d, \| \bullet \|_1) \to (\mathbb{R}, | \bullet |),$$

is continuous. In particular, the restriction of this function is continuous on the following subset of $\mathbb{R}^d$,

$$S_1^{\|\bullet\|_1}(0) := \{\vec{v} \in \mathbb{R}^d : \|\vec{v}\|_1 = 1\}.$$

This subset is closed and bounded. Therefore the continuous function $\| \bullet \|$ takes on a minimal value B on this subset. Since the subset does not contain the zero vector, B is a positive real number.

Of course $\|\vec{0}\|$ and $\|\vec{0}\|_1$ both equal 0, so that $\|\vec{0}\|_1 \leq (1/B)\|\vec{0}\|$. For every nonzero vector $\vec{w}$ in $\mathbb{R}^d$, $\vec{w}$ equals $\|\vec{w}\|_1 \cdot \vec{v}$ for the vector $\vec{v} = \vec{w}/\|\vec{w}\|_1$ that is contained in $S^1_{\|\bullet\|_1}$. By homogeneity of $\| \bullet \|$,

$$\|\vec{w}\|_1 = \frac{\|\vec{w}\|}{\|\vec{v}\|}.$$

Since $\vec{v}$ is in $S^1_{\|\bullet\|_1}$,

$$\|\vec{v}\| \geq B.$$

Therefore,

$$\|\vec{w}\|_1 \leq \frac{\|\vec{w}\|}{B} = (1/B)\|\vec{w}\|.$$

So, since $\| \bullet \|$ is C-Lipschitz with respect to $\| \bullet \|_1$, and since $\| \bullet \|_1$ is $(1/B)$-Lipschitz with respect to $\| \bullet \|$, these are equivalent norms. ☐

Since these norms are equivalent, the corresponding metrics functions are equivalent. Since the metric functions are equivalent, the corresponding topologies on $\mathbb{R}^d$ are equal. The **norm topology** on $\mathbb{R}^d$ is defined to be the unique topology on $\mathbb{R}^d$ induced by any of the (equivalent) norms on $\mathbb{R}^d$. It is also often called the "Euclidean topology", the "analytic topology" (for "analysis"), or the "classical topology". Since every finite dimensional $\mathbb{R}$-vector space V is isomorphic to $\mathbb{R}^d$ for some nonnegative integer d, the same holds for V: any two norms are equivalent, and all norms induce the same topology on V, called the norm topology. For infinite dimensional vector spaces, there are typically many inequivalent norms, e.g., the ℓ_p-norms on $\mathbb{R}^\infty$ for $1 \leq p < \infty$ are all inequivalent,

$$\|(x_1, x_2, x_3, \dots, x_n, 0, \dots, 0, \dots,)\|_p := \sqrt[p]{|x_1|^p + \cdots + |x_n|^p}.$$

Jason Starr

3 Comparison of Notions of Continuity

Let (X, d) and (Y, e) be metric spaces. Let $f : (X, d) \to (Y, e)$ be a continuous function. Let $x \in X$ be an element. As proved, f is continuous at x if and only if for every open neighborhood $V \subset Y$ of $f(x)$, $f^{-1}(V)$ contains an open neighborhood of x. Similarly, f is (everywhere) continuous if and only if for every open subset $V \subset Y$, $f^{-1}(V)$ is an open subset of X. Thus, continuity of f is a "topological property". However, there are several variants of continuity that are useful in analysis that are not strictly topological.

A sequence $(x_n)_{n\in\mathbb{N}}$ of elements $x_n \in X$ is **Cauchy** if for every real $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for every $m, n \in \mathbb{N}$ with $m, n \geq N$, $d(x_n, x_m)$ is less than ϵ. The function f is **Cauchy preserving** if for every Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ in (X, d), also $(f(x_n))_{n\in\mathbb{N}}$ is Caucy in (Y, e).

Lemma 3.1. *Every Cauchy preserving function is continuous. Also, the composition of Cauchy preserving functions is Cauchy preserving.*

Proof. Let $f : (X, d) \to (Y, e)$ be a function that is Cauchy preserving. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in (X, d) that converges (with respect to d) to $x \in X$. Form the new sequence $(\widetilde{x}_m)_{m\in\mathbb{N}}$ by $\widetilde{x}_{2n} = x_n$ and $\widetilde{x}_{2n+1} = x$. Then also $(\widetilde{x}_m)_{m\in\mathbb{N}}$ converges to x, hence it is Cauchy. Since f is Cauchy preserving, also the sequence $(f(\widetilde{x}_m))_{m\in\mathbb{N}}$ is Cauchy (with respect to e). In particular, for every real $\epsilon > 0$, there exists an integer $N \in \mathbb{N}$ such that for every $m, n \geq N$, $e(f(\widetilde{x}_m), f(\widetilde{x}_n)) < \epsilon$. In particular, $f(\widetilde{x}_{2N+1})$ equals $f(x)$. Hence, for every $n \geq N$, since $2n > N$, $e(f(x), f(x_n)) < \epsilon$. Thus $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(x)$. Since this holds for every convergent sequence in (X, d), f is continuous.

Let (T, c) be a metric space, and let $g : (T, c) \to (X, d)$ be a function. Assume that both g and f are Cauchy preserving. Let $(t_n)_{n\in\mathbb{N}}$ be a sequence in (T, c) that is Cauchy. Since g is Cauchy preserving, the sequence $(g(t_n))_{n\in\mathbb{N}}$ is Cauchy in (X, d). Since f is Cauchy preserving, the sequence $(f(g(t_n)))_{n\in\mathbb{N}}$ is Cauchy in (Y, e). This sequence equals $((f \circ g)(t_n))_{n\in\mathbb{N}}$. Since this holds for every Cauchy sequence $(t_n)_{n\in\mathbb{N}}$ in (T, c), the function $f \circ g$ is Cauchy preserving. □

Not every continuous function is Cauchy preserving. For instance, for $X = Y = \{1/n : n \in \mathbb{N}\}$, for d the restriction of the Euclidean metric, and for e the discrete metric, then the identity function $f : (X, d) \to (Y, e)$ is continuous, yet the Cauchy sequence $(1/n)_{n\in\mathbb{N}}$ in (X, d) is mapped to a non-Cauchy sequence in (Y, e).

As studied in other analysis courses, for every metric space (X, d), there exists an isometric (hence injective) function $f : (X, d) \to (\overline{X}, \overline{d})$ such that (i) a sequence $(x_n)_{n\in\mathbb{N}}$ in (X, d) is convergent in $(\overline{X}, \overline{d})$ if and only if $(x_n)_{n\in\mathbb{N}}$ is Cauchy in (X, d), and (ii) every element of $\overline{X}$ is the convergent limit in $(\overline{X}, \overline{d})$ of a sequence $(x_n)_{n\in\mathbb{N}}$ of elements $x_n \in X$. The triple $(\overline{X}, \overline{d}, i)$ is unique up to unique isometry, and it is called the *completion* of (X, d). For the completion $j : (Y, e) \to (\overline{Y}, \overline{e})$, for a continuous function $f : (X, d) \to (Y, e)$, f is Cauchy preserving if and only if there exists a continuous function $\overline{f} : (\overline{X}, \overline{d}) \to (\overline{Y}, \overline{e})$ with $\overline{f} \circ i$ equals $j \circ f$.

A function $f : (X, d) \to (Y, e)$ is **uniformly continuous** if for every real $\epsilon > 0$, there exists $\delta > 0$ such that for every $x, x' \in X$ with $d(x, x') < \delta$, also $e(f(x), f(x')) < \epsilon$.

Lemma 3.2. *Every uniformly continuous function is Cauchy preserving. Also, the composition of uniformly continuous functions is uniformly continuous.*

Proof. Let $f : (X, d) \to (Y, e)$ be a function that is uniformly continuous. Let $(x_n)_{n\in\mathbb{N}}$ be Cauchy. For every real $\epsilon > 0$, since f is uniformly continuous, there exists real $\delta > 0$ such that for every $x, x' \in X$ with $d(x, x') < \delta$, also $e(f(x), f(x')) < \epsilon$. Since $(x_n)_{n\in\mathbb{N}}$ is Cauchy, there exists $N \in \mathbb{N}$ such that for every $m, n \in \mathbb{N}$ with $m, n \geq N$, $d(x_m, x_n) < \delta$. Thus, also $e(f(x_m), f(x_n)) < \epsilon$. Since this holds for every real $\epsilon > 0$, the sequence $(f(x_n))_{n\in\mathbb{N}}$ is Cauchy. Since this holds for every Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ in (X, d), f is Cauchy preserving.

Let (T, c) be a metric space, and let $g : (T, c) \to (X, d)$ be a function. Assume that both g and f are uniformly continuous. Let $\epsilon > 0$ be a real number. Since f is uniformly continuous, there exists real $\delta > 0$ such that for every $x, x' \in X$ with $d(x, x') < \delta$, also $e(f(x), f(x')) < \epsilon$. Since g is uniformly continuous, there exists real $\gamma > 0$ such that for every $t, t' \in T$ with $c(t, t') < \gamma$, also $d(g(t), g(t')) < \delta$. Thus, also $e(f(g(t)), f(g(t'))) < \epsilon$. In other words, if $c(t, t') < \gamma$, then $e((f \circ g)(t), (f \circ g)(t')) < \gamma$. Since this holds for every real $\epsilon > 0$, $f \circ g$ is uniformly continuous. □

Not every Cauchy preserving function is uniformly continuous. For instance, let $X \subset \mathbb{R}$ be the subset $\{2n + (1/m) : m, n \in \mathbb{N}, 1 \leq m \leq n\}$, and let d be the restriction of the Euclidean metric. Let Y equal X, and let e be the discrete metric on Y. Let $f : X \to Y$ be the identity function. Then f is continuous, and f preserves Cauchy sequences in X, since every Cauchy sequence in X is eventually constant. However, f is not uniformly continuous. If the completion $(\overline{X}, \overline{d})$ is compact, then every Cauchy preserving function $f : (X, d) \to (Y, e)$ is uniformly continuous.

A function $f : (X, d) \to (Y, e)$ is C-**Lipschitz** if for every $x, x' \in X$, $e(f(x), f(x')) \leq C \cdot d(x, x')$. If f is C-Lipschitz, then for every real $C' \geq C$, also f is C'-Lipschitz. The function f is **Lipschitz** if there exists some real $C \geq 0$ such that f is Lipschitz (the function is 0-Lipschitz if and only if the function is constant).

Lemma 3.3. *Every Lipschitz function is uniformly continuous. Also, the composition of Lipschitz functions is Lipschitz.*

Proof. Let $f : (X, d) \to (Y, e)$ be a C-Lipschitz function. If C equals 0, then it is also 1-Lipschitz, hence assume that $C > 0$. For every real $\epsilon > 0$, let δ equal ϵ/C, which is a positive real number. For every $x, x' \in X$, if $d(x, x') < \delta$, then $C \cdot d(x, x') < \epsilon$. Thus, since f is C-Lipschitz, $e(f(x), f(x')) < \epsilon$. Thus f is uniformly continuous.

Let (T, c) be a metric space, and let $g : (T, c) \to (X, d)$ be a function. Assume that g is B-Lipschitz and that f is C-Lipschitz for real numbers $C, B \geq 0$. Then for every $t, t' \in T$, since f is C-Lipschitz, $e((f \circ g)(t), (f \circ g)(t')) \leq C \cdot d(g(t), g(t'))$. Since g is B-Lipschitz, $d(g(t), g(t')) \leq B \cdot c(t, t')$. Hence, altogether $f \circ g$ is $B \cdot C$-Lipschitz. □

Not every uniformly continuous function is Lipschitz. The standard example is the identity function $\mathrm{Id} : ([0, 1], d) \to ([0, 1], e)$, where e is the Euclidean metric and $d(x, y) = |x^2 - y^2|$. For every real

$\epsilon > 0$, for every $x, y \in [0,1]$ with $|x^2 - y^2| < epsilon^2$, then $|x-y| < \epsilon$, so that f is uniformly continuous. Yet for every real $C > 0$, for $x = 1/(1+C)$ and for $y = 0$,

$$e(f(x), f(y)) = 1/(1+C) > C \cdot \frac{1}{(1+C)^2} = C \cdot d(x,y).$$

Hence f is not C-Lipschitz.

For a set X, two metric functions d and e on X *induce the same topology* or are **homeomorphic** if both $\mathrm{Id}_X : (X,d) \to (X,e)$ and $\mathrm{Id}_X : (X,e) \to (X,d)$ are continous. This holds if and only if the topology $\mathcal{T}_d$ on X equals the topology $\mathcal{T}_e$ on X. The metric functions are **Cauchy equivalent** if both $\mathrm{Id}_X : (X,d) \to (X,e)$ and $\mathrm{Id}_X : (X,e) \to (X,d)$ are Cauchy preserving. This holds if and only if the identity map on X extends to a unique homeomorphism between the completion of (X,d) and the completions of (X,e). The metric functions are **uniform** with respect to each other if both $\mathrm{Id}_X : (X,d) \to (X,e)$ and $\mathrm{Id}_x : (X,e) \to (X,d)$ are uniformly continuous. This holds, for instance, if d and e are homeomorphic and X is compact for this topology. Finally, as discussed above, the metric functions are **equivalent** if both $\mathrm{Id}_X : (X,d) \to (X,e)$ and $\mathbf{Id}_X : (X,e) \to (X,d)$ are Lipschitz. By the lemmas above, equivalent metrics are uniform, uniform metrics are Cauchy equivalent, and Cauchy equivalent metrics are homeomorphism. As the examples above show, all of these implications are strict.

Because equivalence of metrics implies uniform, Cauchy equivalence, and homomorphic, in testing whether a given $f : (X,d) \to (Y,e)$ is continuous, resp. Cauchy preserving, uniformly continuous, Lipschitz, it suffices to test after replacing the metrics by equivalent metrics. In particular, for a metric space (T,c) and functions $f : (T,c) \to (X,d)$ and $g : (T,c) \to (Y,e)$, to test whether the product function $(f,g) : T \to X \times Y$ is continuous, resp. Cauchy preserving, uniformly continuous, Lipschitz, we may use any of the equivalent product metrics on $X \times Y$.

Lemma 3.4. *The product function (f,g) is continuous, resp. Cauchy preserving, uniformly continuous, Lipschitz, if both f and g are continuous, resp. Cauchy preserving, uniformly continuous, Lipschitz.*

Proof. Use the product metric $\rho((x,y),(x',y')) = d(x,x') + e(y,y')$. With this product metric, all of these implications follow in a straightforward manner. □

4 Bounded Linear Transformations on Normed Vector Spaces

Let $(V, \|\bullet\|_V)$ and $(W, \|\bullet\|_W)$ be normed vector spaces. A linear transformation $T : V \to W$ is **bounded** with respect to $\|\bullet\|_V$ and $\|\bullet\|_W$ if T is continuous with respect to the induced metrics.

Proposition 4.1. *A linear transformation T is bounded if and only if it is Lipschitz.*

Proof. By the lemmas above, every Lipschitz function is continuous. Conversely, assume that T is continuous. In particular, T is continuous at 0. Thus, for $\epsilon = 1$, there exists real $\delta > 0$ such that for every $\mathbf{v} \in V$ with $\|\mathbf{v}\|_V < \delta$, also $\|T(\mathbf{v})\|_W < 1$. For every nonzero $\mathbf{u} \in V$, define

$$\mathbf{v} := \left(\frac{\delta}{2\|\mathbf{u}\|_V}\right)\mathbf{u}.$$

Then $\|\mathbf{v}\|_V$ equals $\delta/2$, so that $\|T(\mathbf{v})\|_W < 1$. Since T is linear and since $\|\bullet\|_W$ is homogeneous,

$$\|T(\mathbf{v})\|_W = \left(\frac{\delta}{2\|\mathbf{u}\|_V}\right)\|T(\mathbf{u})\|_W.$$

Thus, we have

$$\|T(\mathbf{u})\|_W \leq \frac{2}{\delta}\|\mathbf{u}\|_V.$$

This is trivially also true for $\mathbf{u} = 0$.

Now let $\mathbf{v}_1$,$\mathbf{v}_2$ be elements of V. Since T is linear,

$$d_{\|\bullet\|_W}(T(\mathbf{v}_1), T(\mathbf{v}_2)) = \|T(\mathbf{v}_1) - T(\mathbf{v}_2)\|_W = \|T(\mathbf{v}_1 - \mathbf{v}_2)\|_W.$$

Using the inequality above,

$$d_{\|\bullet\|_W}(T(\mathbf{v}_1), T(\mathbf{v}_2)) \leq \frac{2}{\delta}\|\mathbf{v}_1 - \mathbf{v}_2\|_V = \frac{2}{\delta}d_{\|\bullet\|_V}(\mathbf{v}_1, \mathbf{v}_2).$$

Therefore T is Lipschitz with constant $2/\delta$ (in fact, any constant less than $1/\delta$ would have worked). ☐

As is clear from the proof, a linear transformation is bounded if and only if for the unit ball $U_1^{\|\bullet\|_V}(0)$ about 0, the image set $T(U_1^{\|\bullet\|_V}(0))$ is bounded, i.e., there exists a real number $B > 0$ such that $T(U_1^{\|\bullet\|_V}(0)) \subseteq U_B^{\|\bullet\|_W}(0)$. In this case, B is a Lipschitz constant for T. The Lipschitz constants of a bounded linear operator are quite important in later analysis classes. The infimum over all Lipschitz constants is the **operator norm**, $\|T\|_{\text{op}}$ of T with respect to $\|\bullet\|_V$ and $\|\bullet\|_W$, i.e., the smallest nonnegative real number such that for every $\mathbf{v} \in V$,

$$\|T(\mathbf{v})\|_W \leq \|T\|_{\text{op}} \cdot \|\mathbf{v}\|_V.$$

Corollary 4.2. *Let V be a vector space. Norms on V induce the same topology if and only if they are equivalent.*

Proof. Norms on V induce the same topology if and only if the identity function is continuous in both directions for the metrics induced by the norms. But then, by the proposition, there are Lipschitz constants for the identity function in each direction, i.e., the norms are equivalent. ☐

In the finite dimensional case, all norms are equivalent. Even more, all linear transformations on a finite dimensional domain are bounded.

Lemma 4.3. *If V is finite dimensional, every linear transformation $T : (V, \| \bullet \|_V) \to (W, \| \bullet \|_W)$ is bounded.*

Proof. Let $\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ be an ordered basis for V. Since all norms on V are equivalent, replace the norm by the following ℓ_1-norm: for every $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$,

$$\|\mathbf{v}\|_V = \|a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n\|_V = |a_1| + \cdots + |a_n|.$$

Since T is a linear transformation,

$$T(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = a_1T(\mathbf{v}_1) + \cdots + a_nT(\mathbf{v}_n).$$

Define B to be the nonnegative real number,

$$B = \max(\|T(\mathbf{v}_1)\|_W, \dots, \|T(\mathbf{v}_n)\|_W).$$

Then for every $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ in V, by the triangle inequality and homogeneity for $\| \bullet \|_W$

$$\|T(a_1\mathbf{v}_1+\cdots+a_n\mathbf{v}_n)\|_W = \|a_1T(\mathbf{v}_1)+\cdots+a_nT(\mathbf{v}_n)\| \leq |a_1|\|T(\mathbf{v}_1)\|+\cdots+|a_n|\|T(\mathbf{v}_n)\| \leq B\|\mathbf{v}\|_V.$$

Thus T is a bounded linear transformation with $\|T\|_{\text{op}} \leq B$. □

5 Derivatives are Invariant under Equivalent Norms

Let $(V, \| \bullet \|_V)$ and $(W, \| \bullet \|_W)$ be normed vector spaces. Let $U \subset V$ be an open subset, and let $\mathbf{a} \in U$ be an element. Then there exists real $r > 0$ such that the neighborhood $U_r^{\|\bullet\|_V}(\mathbf{a})$ is contained in U. Let $f : U \to W$ be a function. Let $T : V \to W$ be a bounded linear transformation. By definition, the bounded linear transformation T is a **derivative** of f at $\mathbf{a}$ if for every real $\epsilon > 0$, there exists real δ with $0 < \delta < r$ such that for all $\mathbf{v} \in V$ with $\|\mathbf{v}\|_V < \delta$,

$$\|f(\mathbf{a}+\mathbf{v}) - f(\mathbf{a}) - T(\mathbf{v})\|_W \leq \epsilon \cdot \|\mathbf{v}\|_V.$$

The function f is **differentiable** at $\mathbf{a}$ if there exists a bounded linear transformation T that is a derivative of f at $\mathbf{a}$.

Lemma 5.1. *If a derivative of f at $\mathbf{a}$ exists, then it is unique.*

Proof. Let $S, T : V \to W$ be linear transformations that are derivatives of f at $\mathbf{a}$. The key observation is that for all $\mathbf{v} \in V$ with $\|\mathbf{v}\|_V < r$,

$$S(\mathbf{v}) - T(\mathbf{v}) = (f(\mathbf{a}+\mathbf{v}) - f(\mathbf{a}) - T(\mathbf{v})) - (f(\mathbf{a}+\mathbf{v}) - f(\mathbf{a}) - S(\mathbf{v})).$$

For every real $\epsilon > 0$, there exists a δ with $0 < \delta < r$ that simultaneously satisfies the derivative inequality for both S and T. For every nonzero vector $\mathbf{u} \in V$, defining $\mathbf{v} = \delta\mathbf{u}/(2\|\mathbf{u}\|_V)$ as above, then

$$\|T(\mathbf{u}) - S(\mathbf{u})\|_W = \frac{2\|\mathbf{u}\|_V}{\delta} \cdot \|T(\mathbf{v}) - S(\mathbf{v})\| \leq \frac{2\|\mathbf{u}\|_V}{\delta} \cdot \epsilon \cdot (\|\mathbf{v}\|_V + \|\mathbf{v}\|_V) = 2\epsilon\|\mathbf{u}\|_V.$$

Since this holds for every real $\epsilon > 0$, $\|T(\mathbf{u}) - S(\mathbf{u})\|_W$ equals 0. Since $\| \bullet \|_W$ is positive definite, $T(\mathbf{u})$ equals $S(\mathbf{u})$ for every nonzero $\mathbf{u}$ in V. Since S and T are linear, also $T(0) = 0 = S(0)$. Thus S equals T. □

In fact, the derivative is a topological property for normed vector spaces. Since homeomorphic norms are equivalent, it suffices to work directly with equivalent norms. Let $| \bullet |_V$ be a norm that is equivalent to $\| \bullet \|_V$, i.e., there exists $c \geq 1$ such that for every $\mathbf{v} \in V$,

$$\|\mathbf{v}\|_V \leq c \cdot |\mathbf{v}|_V, \ \ |\mathbf{v}|_V \leq c\|\mathbf{v}\|_V.$$

Similarly, let $| \bullet |_W$ be a norm that is equivalent to $\| \bullet \|_W$, i.e., there exists $b \geq 1$ such that for every $\mathbf{v} \in W$,

$$\|\mathbf{w}\|_W \leq a \cdot |\mathbf{w}|_W, \ \ |\mathbf{w}|_W \leq a\|\mathbf{w}\|_W.$$

Lemma 5.2. *The function $f : U \to W$ is differentiable at $\mathbf{a}$ with respect to $\| \bullet \|_V$ and $\| \bullet \|_W$ if and only if f is differentiable at $\mathbf{a}$ with respect to $| \bullet |_V$ and $| \bullet |_W$. In this case, the derivative is independent of the choice of (equivalent) norm.*

Proof. Assume that f is differentiable at $\mathbf{a}$ with derivative T with respect to $\| \bullet \|_V$ and $\| \bullet \|_W$. Then for every real $\epsilon > 0$, set $\epsilon' = \epsilon/(bc)$ which is still positive. There exists real δ' with $0 < \delta' < r$ such that for every $\mathbf{v} \in V$ with $\|\mathbf{v}\|_V < \delta'$,

$$\|f(\mathbf{a} + \mathbf{v}) - f(\mathbf{a}) - T(\mathbf{v})\|_W \leq \epsilon' \cdot \|\mathbf{v}\|_V.$$

Thus,

$$|f(\mathbf{a} + \mathbf{v}) - f(\mathbf{a}) - T(\mathbf{v})|_W \leq c \cdot \|f(\mathbf{a} + \mathbf{v}) - f(\mathbf{a}) - T(\mathbf{v})\|_W \leq c\epsilon \cdot \|\mathbf{v}\|_V \leq c\epsilon \cdot b|\mathbf{v}\ |_V = \epsilon \cdot \ | \ \mathbf{v}|_V.$$

Finally, if $|\mathbf{v}|_V < \delta'/b$, then $\|\mathbf{v}\|_V < \delta'$, so that the inequality above holds. Thus, setting $\delta = \delta'/b$, f is differentiable at $\mathbf{a}$ with derivative T with respect to $| \bullet |_V$ and $| \bullet |_W$. □

6 The Critical Point Theorem and the Mean Value Theorem

In the proof of the Inverse Function Theorem, we used the multivariable version of the Critical Point Theorem. Let $(V, \| \bullet \|_V)$ be a normed vector space. Let $U \subset V$ be an open subset. Let $h : U \to \mathbb{R}$ be a function.

Theorem 6.1 (Critical Point Theorem)**.** *Let $\mathbf{c} \in U$ be an element such that $h(\mathbf{c})$ is a minimum of $h(U)$ (or a maximum). For each nonzero vector $\mathbf{u} \in V$ for which the directional derivative $h'(\mathbf{c}; \mathbf{u})$ is defined, $h'(\mathbf{c}; \mathbf{u})$ equals 0.*

Proof. Because U is open, there exists a positive real r such that for every $t \in (-r, +r)$, $\mathbf{c} + t\mathbf{u}$ is in U. Consider the single-variable, real-valued function $g : (-r, r) \to \mathbb{R}$ defined by $g(t) = h(\mathbf{c} + t\mathbf{u})$. By hypothesis, $t = 0$ is a minimum. Also, by definition, the derivative $g'(0)$ equals the directional derivative $h'(\mathbf{c}; \mathbf{u})$. Thus, by the single-variable Critical Point Theorem, if $h'(\mathbf{c}; \mathbf{u})$ is defined, then $h'(\mathbf{c}; \mathbf{u})$ equals 0. □

Let $\mathbf{a}, \mathbf{b} \in U$ be distinct points such that the closed line segment $L = \{(1-t)\mathbf{a} + t\mathbf{b} : t \in [0, 1]\}$ is contained in U. Denote by L^* the open line segment $L \setminus \{\mathbf{a}, \mathbf{b}\}$. Denote $\mathbf{b} - \mathbf{a}$ by $\mathbf{u}$, so that the L is also $\{\mathbf{a} + t\mathbf{u} : t \in [0, 1]\}$.

Theorem 6.2 (Mean Value Theorem)**.** *If the directional derivative $h'(\mathbf{c}; \mathbf{u})$ is defined for every $\mathbf{c} \in L*$ and if $h|L$ is continuous at $\mathbf{a}$ and $\mathbf{b}$, then there exists $\mathbf{c} \in L^*$ such that $h(\mathbf{b}) - h(\mathbf{a})$ equals $h'(\mathbf{c}, \mathbf{u})$.*

Proof. For $t \in \mathbb{R}$, consider the function

$$g(t) = [h(\mathbf{a} + t\mathbf{u}) - h(\mathbf{a})] - t[h(\mathbf{b}) - h(\mathbf{a})].$$

By hypothesis, g is defined on an open neighborhood of $[0, 1]$, g is differentiable at every point of $(0, 1)$, and g is continuous at 0 and 1. By the Chain Rule,

$$g'(t) = h'(\mathbf{a} + t\mathbf{u}; \mathbf{u}) - [h(\mathbf{b}) - h(\mathbf{a})].$$

The theorem is equivalent to the statement that for some $t \in (0, 1)$, $g'(t)$ equals 0. If g is constant, then $g'(t)$ equals 0 for every $t \in (0, 1)$. Thus, assume that g is not constant.

Since g is differentiable at every point of $(0, 1)$, g is continuous at every point of $[0, 1]$. Thus, by the Extreme Value Theorem, g has a minimum value and a maximum value. Since g is not constant, these values are distinct; in particular, they do not both equal 0. By construction $g(0)$ and $g(1)$ both equal 0. Thus there exists $t \in (0, 1)$ such that either $g(t)$ is a minimum or a maximum. By the Critical Point Theorem, $g'(t)$ equals 0. □

For a vector-valued function, e.g., $f : \mathbb{R} \to \mathbb{R}^3$ by $f(t) = (t, t^2, t^3)$, it can easily happen that for no $\mathbf{c} \in L$ is it true that $f(\mathbf{b}) - f(\mathbf{a})$ equals $f'(\mathbf{c}, \mathbf{u})$, e.g., for the function above, this is the case for $\mathbf{a} = 0$ and $\mathbf{b} = 1$. However, quite often we only need an estimate on norms that would be implied by such an extension of the Mean Value Theorem.

Let $(W, \langle \bullet, \bullet \rangle)$ be an inner product space, and let $\| \bullet \|_W$ denote the norm associated to this inner product. Let $f : U \to W$ be a function.

Proposition 6.3 (Mean Value Theorem Estimate). *With notation as above, if the directional derivative $f'(\mathbf{c};\mathbf{u})$ is defined for every $\mathbf{c} \in L^*$ and if $f|L$ is continuous at $\mathbf{a}$ and $\mathbf{b}$, then there exists $\mathbf{c} \in L^*$ such that $\|f(\mathbf{b}) - f(\mathbf{a})\|_W^2 = \langle f'(\mathbf{c};\mathbf{u}), f(\mathbf{b}) - f(\mathbf{a})\rangle$. In particular, $\|f(\mathbf{b}) - f(\mathbf{a})\|_W \leq \|f'(\mathbf{c};\mathbf{u})\|_W$.*

Proof. Denote $\mathbf{w} = f(\mathbf{b}) - f(\mathbf{a})$. Define the function $h : [0,1] \to \mathbb{R}$ by,

$$h(t) = \langle f(\mathbf{a} + t\mathbf{u}) - f(\mathbf{a}), \langle \mathbf{w}\rangle.$$

By the hypotheses and the Chain Rule, h is differentiable for every $t \in (0,1)$ and h is continuous at 0 and 1. By the Chain Rule,

$$h'(t) = \langle f'(\mathbf{a} + t\mathbf{u}; \mathbf{u}), \mathbf{w}\rangle.$$

By the Mean Value Theorem, there exists $t \in (0,1)$ such that $h(1) - h(0)$ equals $h'(t)$, i.e.,

$$\|f(\mathbf{b}) - f(\mathbf{a})\|_W^2 = \langle f'(\mathbf{c};\mathbf{u}), \mathbf{w}\rangle,$$

where $\mathbf{c} = \mathbf{a} + t\mathbf{u}$. From the Cauchy-Schwarz inequality,

$$\|f(\mathbf{b}) - f(\mathbf{a})\|_W^2 \leq \|f'(\mathbf{c};\mathbf{u})\|_W \cdot \|f(\mathbf{b}) - f(\mathbf{a})\|_W.$$

Simplifying, this gives

$$\|f(\mathbf{b}) - f(\mathbf{a})\|_W \leq \|f'(\mathbf{c};\mathbf{u})\|_W.$$

□

This leads to other estimates. For instance, assume that also the directional derivative $f'(\mathbf{a};\mathbf{u})$ is defined.

Corollary 6.4. *With hypotheses as above, assume that also $f'(\mathbf{a};\mathbf{u})$ is defined. Then there exists $\mathbf{c} \in L^*$ such that $\|f(\mathbf{b}) - f(\mathbf{a}) - f'(\mathbf{a};\mathbf{u})\|_W \leq \|f'(\mathbf{c};\mathbf{u}) - f'(\mathbf{a};\mathbf{u})\|_W$.*

Proof. The proof is almost identical to the previous proof, but replacing $\mathbf{w} = f(\mathbf{b}) - f(\mathbf{a}) - f'(\mathbf{a};\mathbf{u})$ and replacing

$$h(t) = f(\mathbf{a} + t\mathbf{u}) - f(\mathbf{a}) - tf'(\mathbf{a};\mathbf{u}).$$

□

7 Injective Derivative Implies Local Injectivity

The proof in the book of Lemma 8.1 is correct and complete, but the estimates involve the dimensions is an unnecessary way. As in the previous section, let $(V, \|\bullet\|_V)$ be a normed vector space, let $(W, \langle \bullet, \bullet \rangle)$ be an inner product space, let $U \subset V$ be an open subset, let $f : U \to W$ be a function, and let $\mathbf{a} \in U$ be an element. Assume that for every $\mathbf{b} \in U$, f is differentiable at $\mathbf{b}$. Moreover, assume that the function $\mathbf{b} \mapsto Df_{\mathbf{b}}$ is continuous at $\mathbf{a}$.

Lemma 7.1. *Assume that there exists $\alpha > 0$ such that for every $\mathbf{u} \in U$, $\|Df_{\mathbf{a}}(\mathbf{u})\|_W \geq \alpha \|\mathbf{u}\|_V$. Then for every real β with $0 < \beta < \alpha$, there exists $\delta > 0$ such that for every $\mathbf{x}, \mathbf{y} \in B_\delta(\mathbf{a})$, $\|f(\mathbf{x}) - f(\mathbf{y})\|_W \geq \beta \|\mathbf{x} - \mathbf{y}\|_V$.*

Proof. Of course the inequality is trivially satisfied if $\mathbf{x}$ equals $\mathbf{y}$. Thus assume that $\mathbf{x}$ and $\mathbf{y}$ are distinct. Also, by the triangle inequality, for every pair $\mathbf{x}$, $\mathbf{y}$ of distinct elements of $B_\delta(\mathbf{a})$, also the line segment $L = \{t\mathbf{x} + (1-t)\mathbf{y} : t \in [0,1]\}$ is contained in $B_\delta(\mathbf{a})$.

Define $\epsilon = \alpha - \beta$, which is positive by hypothesis. Since $Df_{\mathbf{b}}$ is continuous at $\mathbf{a}$, there exists positive real δ such that $B_\delta(\mathbf{a})$ is contained in U and for every $\mathbf{c} \in B_\delta(\mathbf{a})$, $\|Df_{\mathbf{b}} - Df_{\mathbf{a}}\|_{\text{op}} < \epsilon$, i.e., for every $\mathbf{u} \in V$,

$$\|Df_{\mathbf{b}}(\mathbf{u}) - Df_{\mathbf{a}}(\mathbf{u})\|_W \leq \epsilon \|\mathbf{u}\|_V.$$

Consider the modified function, $H : U \to W$ defined by

$$H(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{a}) - Df_{\mathbf{a}}(\mathbf{x} - \mathbf{a}).$$

This function measures the deviation of $f(\mathbf{x})$ from the "best approximation" affine linear function $f(\mathbf{a}) + Df_{\mathbf{a}}(\mathbf{x} - \mathbf{a})$ at $\mathbf{a}$. In particular, for every $\mathbf{b} \in B_\delta(\mathbf{a})$,

$$DH_{\mathbf{b}} = Df_{\mathbf{b}} - Df_{\mathbf{a}}.$$

Thus, for every $\mathbf{u} \in V$, there is an equation of directional derivatives,

$$H'(\mathbf{b}; \mathbf{u}) = (Df_{\mathbf{b}} - Df_{\mathbf{a}})(\mathbf{u}).$$

Also, for every $\mathbf{x}, \mathbf{y} \in B_\delta(\mathbf{a})$,

$$H(\mathbf{x}) - H(\mathbf{y}) = [f(\mathbf{x}) - f(\mathbf{y})] - Df_{\mathbf{a}}(\mathbf{x} - \mathbf{y}).$$

By the triangle inequality,

$$\|Df_{\mathbf{a}}(\mathbf{x} - \mathbf{y})\|_W \leq \|H(\mathbf{x}) - H(\mathbf{y})\|_W + \|f(\mathbf{x}) - f(\mathbf{y})\|_W,$$

or equivalently,

$$\|f(\mathbf{x}) - f(\mathbf{y})\|_W \geq \|Df_{\mathbf{a}}(\mathbf{x} - \mathbf{y})\|_W - \|H(\mathbf{x}) - H(\mathbf{y})\|_W.$$

Setting $\mathbf{u} = \mathbf{x} - \mathbf{y}$, by the Mean Value Theorem Estimate there exists $\mathbf{c} \in L^*$ such that

$$\|H(\mathbf{x}) - H(\mathbf{y})\|_W \leq \|H'(\mathbf{c}; \mathbf{u})\|_W = \|(Df_{\mathbf{c}} - Df_{\mathbf{b}})(\mathbf{u})\|_W \leq \epsilon \|\mathbf{u}\|_V.$$

Thus,

$$\|f(\mathbf{x}) - f(\mathbf{y})\|_W \geq \|Df_{\mathbf{a}}(\mathbf{x} - \mathbf{y})\|_W - \epsilon \|_{\|}\mathbf{x} - \mathbf{y}_V.$$

Finally, by the hypothesis on $Df_{\mathbf{a}}$ and α, this gives,

$$\|f(\mathbf{x}) - f(\mathbf{y})\|_W \geq \alpha \|\mathbf{x} - \mathbf{y}\|_V - \epsilon \|\mathbf{x} - \mathbf{y}\|_V \geq \beta \|\mathbf{x} - \mathbf{y}\|_V.$$

$\square$