II. Preliminaries: Advanced Calculus

Topology

**** continuity ****

Recall that

$$f : \mathbb{R} \to \mathbb{R} \text{ is continuous at } t$$

means that

$$\lim_{n\to\infty} x_n = t \quad \Longrightarrow \quad \lim_{n\to\infty} f(x_n) = f(t).$$

Equivalently, it means that

$$\forall\{\varepsilon > 0\}\exists\{w > 0\}\ \ |x - t| < w \quad \Longrightarrow \quad |f(x) - f(t)| < \varepsilon.$$

The abstract (yet eminently practical) intuitive idea behind this is the following. Suppose $f : T \to U$, and we want to know $u := f(t)$. Usually, we do not know t exactly, we only know that t lies in some 'small' set N, and must wonder just how 'small' the corresponding set $f(N) := \{f(s) : s \in N\}$ is, i.e., how 'closely' we may know $f(t)$ in this case. More than that, we would like to be able to make $f(N)$ 'small' by making N suitably 'small'. We can do just that in case f is 'continuous' at t in the following intuitive sense: For every 'neighborhood' M, however 'small', of the point $f(t) \in U$, there is some 'neighborhood' N of $t \in T$ that gets entirely mapped into M by f.

Of course, this requires some reasonable definition of 'neighborhood' of a point, and this leads to one of several equivalent ways to define a 'topology'. For its discussion, the following notations are useful.

**** working with a collection of subsets ****

For $\mathbf{A}$, $\mathbf{C}$ collections of subsets of some set T and f some map on T, we abbreviate

$$\mathbf{A}\Cap\mathbf{C} := \{A \cap C : A \in \mathbf{A}, C \in \mathbf{C}\}, \qquad f\mathbf{A} := \{f(A) : A \in \mathbf{A}\},$$

and use

(1) Definition. $\mathbf{A} \succ \mathbf{C}$ *(in words:* **A refines C***)* $:= \quad \forall\{C \in \mathbf{C}\}\ \exists\{A \in \mathbf{A}\}\ A \subseteq C.$

Thus, $\mathbf{A} \supseteq \mathbf{C} \quad \Longrightarrow \quad \mathbf{A} \succ \mathbf{C}$, but the converse does not hold. In fact, it may be possible to have $\mathbf{A} \succ \mathbf{C}$ even though $\mathbf{A}$ is much smaller than $\mathbf{C}$ in some sense, e.g., even though $\mathbf{A} \subset \mathbf{C}$. Note that $\mathbf{A} \succ \mathbf{C}$ implies that $f\mathbf{A} \succ f\mathbf{C}$.

**** topology defined ****

We specify a topology (i.e., a way of telling far from near) in T by associating with each $t \in T$ a **neighborhood system** $\mathbf{B}(t)$, i.e., a nonempty collection of sets, all containing t. For this to work out properly, this specification has to satisfy two technical assumptions. The first is that the intersection of any two neighborhoods of t contains a neighborhood of t:

(2) Neighborhood Assumption 1. $\forall\{t \in T\}\ \mathbf{B}(t) \succ \mathbf{B}(t) \Cap \mathbf{B}(t)$.

The second is given in (5) below. Three instructive *examples* are $\mathbf{B}(t) := \{\{t\}\}$, $\mathbf{B}(t) := \{T\}$, and $\mathbf{B}(t) := \{S \subset T : t \in S, \#(T\backslash S) < \infty\}$. For $T = \mathbb{R}^n$, a standard choice is $\mathbf{B}(t) := \{B_\varepsilon(t) : \varepsilon > 0\}$, with $B_r(t) := \{s \in T : |s - t| < r\}$ the **open ball around** t **of radius** r, and $|a| := \sqrt{a^t a} = \big(\sum_i |a(i)|^2\big)^{1/2}$ the **Euclidean** length of the n-vector a. These are good examples to try out on the general material to follow.

With the specification of a (proper) neighborhood system $\mathbf{B}(t)$ for each t, the set T, or more precisely, the pair $(T, \mathbf{B})$ becomes a **topological space** (=: **ts**).

(3) Definition. *T, U ts's, $f : T \to U$. f is* **continuous** *at $t \in T$ $:=$ $f\mathbf{B}(t) \succ \mathbf{B}(f(t))$.*

For *example*, if $T = \mathbb{R} = U$ and $\mathbf{B}(t) = \{B_\varepsilon(t) : \varepsilon > 0\}$, with $B_\varepsilon(t) = (t - \varepsilon \,..\, t + \varepsilon)$, then this definition says that $f : \mathbb{R} \to \mathbb{R}$ is continuous at t iff, for every positive ε, one can find a positive w so that $f(B_w(t)) \subseteq B_\varepsilon(f(t))$. This is exactly the definition recalled earlier.

We say that $f : T \to U$ **is continuous** and write this as

$$f \in C(T \to U)$$

if f is continuous at every $t \in T$.

The identity map $1 : T \to T : t \mapsto t$ is always continuous. Also, the composition of continuous maps is continuous. I.e., if $f : T \to U$ is continuous at t and $g : U \to W$ is continuous at $f(t)$, then $gf : T \to W : t \mapsto g(f(t))$ is continuous at t. This is so since $f\mathbf{B}(t) \succ \mathbf{B}(f(t))$ implies that $g(f\mathbf{B}(t)) \succ g\mathbf{B}(f(t))$, and so, since $g\mathbf{B}(f(t)) \succ \mathbf{B}(gf(t))$, we also have $(gf)\mathbf{B}(t) \succ \mathbf{B}(gf(t))$.

**** equivalent topologies ****

The notion of continuity depends, of course, on the particular choice of topologies or neighborhood systems. For *example*, if we use the **discrete topology** on T, i.e., $\mathbf{B}(t) = \{\{t\}\}$ for all $t \in T$, then every f on T is continuous, while, with the **indiscrete (or, trivial) topology**, i.e., with $\mathbf{B}(t) = \{T\}$, all $t \in T$, only constant functions on T are, offhand, continuous. In general, the fewer sets there are in $\mathbf{B}(f(t))$ and/or the more sets there are in $\mathbf{B}(t)$, the easier it is to prove that f is continuous at t. Yet, even if an alternative neighborhood system $\mathbf{B}'(t)$ is obtained from $\mathbf{B}(t)$ by throwing out some sets, it may still happen that $\mathbf{B}'(t) \succ \mathbf{B}(t)$, i.e., that the identity map

$$1 : T' \to T : t \mapsto t$$

is continuous at t. Here, I have used T' as an abbreviation for $(T, \mathbf{B}')$, i.e., to denote the ts made up of the set T with the alternative neighborhood systems $\mathbf{B}'$.

If $\mathbf{B}'(t) \succ \mathbf{B}(t)$ for all $t \in T$, we write $\mathbf{B}' \succ \mathbf{B}$ and say that (the topology provided by) $\mathbf{B}'$ **is stronger than** (the topology provided by) $\mathbf{B}$. $\mathbf{B}'$ is stronger than $\mathbf{B}$ iff $1 \in C(T' \to T)$, and, in that case, for any $f \in C(T \to U)$, the composite function $T' \xrightarrow{1} T \xrightarrow{f} U$ is also continuous, i.e., $f \in C(T' \to U)$. It follows that $\mathbf{B}$ *and* $\mathbf{B}'$ *lead to the same continuous functions, i.e., $C(T' \to U) = C(T' \to U)$ for all ts U, if and only if* $\mathbf{B}$ *and* $\mathbf{B}'$ *are* **equivalent** in the sense that

$$\mathbf{B}' \succ \mathbf{B} \succ \mathbf{B}' \qquad (=: \mathbf{B} \sim \mathbf{B}'),$$

and this, in turn, holds if and only if both $1 : T \to T'$ and $1 : T' \to T$ are continuous.

This is a very useful concept, because it may lead us to a particularly simple specification of a given topology, and may simplify the determination of whether a given function is continuous.

H.P.(1) Prove: *If* $\mathbf{B}(t) \succ \mathbf{A}$, *then* $\mathbf{B}(t)$ *and* $\mathbf{B}'(t) := \mathbf{B}(t) \cup \mathbf{A}$ *are equivalent.*

For *example*, on $T = \mathbb{R}^n$, we might use the much smaller nbhdsystem

$$\mathbf{B}'(t) := \{B_{1/m}(t) : m = 1, 2, 3, \ldots\}$$

which consists of only countably many sets, yet gives rise to the same notion of continuity as the much richer standard system

$$\mathbf{B}(t) := \{B_\varepsilon(t) : \varepsilon > 0\}$$

or the even richer system

$$\{N \subseteq \mathbb{R}^n : \mathbf{B}(t) \succ \{N\}\}.$$

For *example*, on $T = \mathbb{R}^n$, we might use $\mathbf{B}_\infty(t) := \{B_{\varepsilon,\infty}(t) : \varepsilon > 0\}$, made up of the boxes

$$B_{r,\infty}(t) := \{s \in \mathbb{R}^n : \max_i |s(i) - t(i)| < r\},$$

instead of $\mathbf{B}(t)$. The two topologies are equivalent since each euclidean ball $B_r(t)$ around t is contained in the corresponding box $B_{r,\infty}(t)$, (thus $\mathbf{B}(t) \succ \mathbf{B}_\infty(t)$), while, also, each such euclidean ball contains some box $B_{s,\infty}(t)$ for some appropriately small but positive s (hence also $\mathbf{B}_\infty(t) \succ \mathbf{B}(t)$).

H.P.(2) What happens with this last example as $n \to \infty$?

Thus, if also U' is the same set as U but with an equivalent topology $\mathbf{B}'$, then any function $f : T \to U$ continuous in the original topologies is also continuous with respect to the equivalent topologies since it can be viewed as the composition

$$T' \xrightarrow{1} T \xrightarrow{f} U \xrightarrow{1} U'$$

of continuous maps.

(4) Example Here is an example of two familiar nbhdsystems that are *not* equivalent. The standard nbhdsystem $\mathbf{B}_\infty(f)$ for $f \in \mathbb{R}^S$ associated with '**uniform convergence**' consists of the 'balls'

$$B_r(f) := \{g \in \mathbb{R}^S : \sup_{s \in S} |g(s) - f(s)| < r\}, \qquad r > 0.$$

An alternative nbhdsystem $\mathbf{B}_{pw}(f)$ is associated with '**pointwise convergence**' and consists of the 'balls'

$$B_{r,R}(f) := \{g \in \mathbb{R}^S : \max_{s \in R} |g(s) - f(s)| < r\}, \qquad r > 0, \quad R \subset S, \ \#R < \infty.$$

H.P.(3) (a) Prove that $\mathbf{B}_\infty(f) \succ \mathbf{B}_{pw}(f)$ but that $\mathbf{B}_{pw}(f) \not\succ \mathbf{B}_\infty(f)$ in case $\#S \not< \infty$. (b) Try to explain the terms "uniform" and "pointwise" used here in terms of convergence of a sequence (f_n) in $\mathbb{R}^S$.

** open and closed sets **

Let T, or, more precisely, $(T, \mathbf{B})$, be a ts. Here is a list of standard terminology concerning the possible relationships between a point $t \in T$ and a subset U of T. For it, $\mathbf{B}(t) \Cap U := \mathbf{B} \Cap \{U\}$.

t is an **interior point of** U $:=$ $t \in U^o$ $:=$ $\mathbf{B}(t) \succ \{U\}$, i.e., there exists $N \in \mathbf{B}(t)$ s.t. $N \subseteq U$. E.g., $\forall\{N \in \mathbf{B}(t)\}\ t \in N^o$ trivially.

t is a **closure point of** U $:=$ $t \in U^-$ $:=$ $\{\} \notin \mathbf{B}(t) \Cap U$, i.e., $\forall\{N \in \mathbf{B}(t)\}\ N \Cap U \neq \{\}$.

t is an **isolated point of** U $:=$ $\{t\} \in \mathbf{B}(t) \Cap U$, i.e., $\exists\{N \in \mathbf{B}(t)\}\ N \cap U = \{t\}$.

t is a **cluster point of** U $:=$ $\{\} \notin \mathbf{B}(t) \Cap (U \backslash t)$, i.e., $\forall\{N \in \mathbf{B}(t)\}\ (N \cap U)\backslash t \neq \{\}$, i.e., t is a non-isolated closure point of U.

The set U^o of all interior points of U is called the **interior** of U. The set U^- of all closure points of U is called the **closure** of U, and U is **dense (in T)** in case $U^- = T$. Finally, the set $U^- \backslash U^o =: \partial U$ is the **boundary** of U. We have

$$W \subset U \quad \Longrightarrow \quad W^o \subset U^o, W^- \subset U^-.$$

Also,

$$U^o \subseteq U \subseteq U^-,$$

and equality here deserves a special name:

U is **open** $:=$ $U^o = U$. E.g., T and $\{\}$ are open.

U is **closed** $:=$ $U = U^-$. E.g., T and $\{\}$ are closed.

Note that t is a closure point of U iff t is *not* an interior point of the **complement**

$$\backslash U \ := \ T \backslash U = \{s \in T : s \notin U\}$$

of U (in T). In other words, any subset U of T divides all of T into two disjoint classes: the closure points of U and the interior points of $\backslash U$. I.e., $\backslash(U^-) = (\backslash U)^o$. In particular, *$U$ is closed iff $\backslash U$ is open.*

You should verify that *all the properties just introduced will be unchanged if* $\mathbf{B}$ *is replaced by an equivalent neighborhood system.*

Open sets often provide the most efficient way to describe a topological situation.

For example, the other technical assumption a proper neighborhood system has to satisfy is most easily stated in terms of open sets:

(5) Neighborhood Assumption 2. *For every $t \in T$, the collection*

$$\mathbf{B}^o(t) := \{O \subseteq T : t \in O = O^o\}$$

refines $\mathbf{B}(t)$, i.e., every neighborhood of t contains an open set containing t.

H.P.(4) Prove that (5) is not equivalent to the earlier observation that $\forall\{N \in \mathbf{B}(t)\}\ t \in N^o$. (Hint: Show by an example (e.g., $T := \{1,2,3\}$, $\mathbf{B}(1) := \{\{1,2\}\}$, $\mathbf{B}(2) := \{\{2,3\}\}$, $\mathbf{B}(3) := \{\{3,1\}\}$) that, with only (2) and not (5), the interior of a set need not be open.)

H.P.(5) Prove that the neighborhood systems $\mathbf{B}_\infty$ and $\mathbf{B}_{pw}$ of (4)Example both indeed satisfy (5).

H.P.(6) Prove that (5) is equivalent to the following property (which does not explicitly mention open sets but makes more explicit that (5) is an assumption on the way in which nbhdsystems for different points interact): $\forall\{t \in T,\ N \in \mathbf{B}(t)\}\ \exists\{M \in \mathbf{B}(t)\}\ \forall\{m \in M\}\ \exists\{N_m \in \mathbf{B}(m)\}\ N_m \subseteq N$.

H.P.(7) Prove: *If* $(T, \mathbf{B})$ *is ts, and* $\mathbf{C} : T \to 2^{(2^T)}$ *is equivalent to* $\mathbf{B}$ *(i.e.,* $\forall\{t \in T\}\ \mathbf{B}(t) \succ \mathbf{C}(t) \succ \mathbf{B}(t)$*), then* $\mathbf{C}$ *is a neighborhood system provided that* $\forall\{t \in T, C \in \mathbf{C}(t)\}\ t \in C$.

H.P.(8) Prove: (a) *The closure* U^- *of a set* U *is closed.* (That's good.) Equivalently, *the interior* U^o *of a set* U *is open.* (b) *The collection of open sets is closed under finite intersection and arbitrary union*; i.e., the intersection of finitely many open sets is open, and the union of an arbitrary collection of open sets is open.

(6) Proposition. $\mathbf{B}^o(t)$ *is equivalent to* $\mathbf{B}(t)$.

Proof: (5) explicitly states that $\mathbf{B}^o(t) \succ \mathbf{B}(t)$ while $\mathbf{B}(t) \succ \mathbf{B}^o(t)$ follows from the fact that $\mathbf{B}^o(t)$ consists of open sets containing t. □

Thus, when convenient, we may work with the nbhdsystem $\mathbf{B}^o$. For example, we may take it for granted that any open set is a neighborhood for all its elements.

Further, we are able to prove the following efficient

(7) Characterization of continuity. $f : T \to U$ *is continuous iff* $f^{-1}(O)$ *is open for every open* $O \subseteq U$.

Proof. '$\Longrightarrow$': If O is open and $t \in f^{-1}(O)$, then $O \in \mathbf{B}^o(f(t)) \prec f\mathbf{B}(t)$, therefore $f^{-1}(O) \prec \mathbf{B}(t)$, thus $f^{-1}(O)$ is open.

'$\Longleftarrow$': $\forall\{t\}\ \forall\{O \in \mathbf{B}^o(f(t))\}$ we have $f^{-1}(O) \in \mathbf{B}^o(t)$, so $f\mathbf{B}^o(t) \supseteq \mathbf{B}^o(f(t))$, therefore $f\mathbf{B}^o(t) \succ \mathbf{B}^o(f(t))$, , i.e., f is continuous. □

This characterization of continuity doesn't show the practical purpose of continuity, but is often quite useful and efficient in mathematical arguments.

Here is a typical application of continuity.

(8) Proposition. *If* $f \in C(T \to \mathbb{R})$ *and* $Z \subseteq T$, *then* $\sup f(Z) = \sup f(Z^-)$.

Proof: Recall that, for $R \subseteq \mathbb{R}$, $\sup R := \min\{s \in \mathbb{R} : \forall\{r \in R\}\ r \le s\}$, hence $R \subseteq S \Longrightarrow \sup R \le \sup S$, and $\sup R = \sup R^-$. If f is continuous, then $f^{-1}(f(Z)^-)$ is closed and contains Z, hence contains Z^-. Therefore, $f(Z) \subseteq f(Z^-) \subseteq f(Z)^-$, thus $\sup f(Z)^- = \sup f(Z) \le \sup f(Z^-) \le \sup f(Z)^-$, hence $\sup f(Z) = \sup f(Z^-)$. □

Metric space

** metric **

The standard topology on $\mathbb{R}^n$ is given by a metric.

A **metric space** (=: **ms**) (X, d) is a set X and a **metric** d, i.e., a map $d : X \times X \to \mathbb{R}$ satisfying the three conditions

$d(x,y) \ge 0$, with equality iff $x = y$	**(positive definite)**
$d(x,y) = d(y,x)$	**(symmetry)**
$d(x,y) \le d(x,z) + d(z,y)$	**(triangle inequality)**

Note that the triangle inequality can also be stated as

$$d(x, y) - d(y, z) \le d(x, z),$$

and, by interchanging z and x here, we get the symmetric statement

$$|d(x, y) - d(y, z)| \le d(x, z). \tag{9}$$

Standard *examples* include

$$X = \mathbb{R}^n, \quad d(x, y) := \|x - y\|_p$$

with

$$\|x\|_p := \Big(\sum_{i=1}^n |x(i)|^p\Big)^{1/p}$$

for $1 \le p \le \infty$. The interpretation for $p = \infty$ is that of a limit as $p \to \infty$. Thus

$$\|x\|_\infty := \max_i |x(i)|.$$

For $p = 2$, we get the 'usual' metric of $\mathbb{R}^n$, the **Euclidean** metric.

By letting here $n \to \infty$, one obtains the ms $\ell_p := \{x \in \mathbb{R}^{\mathbb{N}} : \|x\|_p < \infty\}$ with metric $d(x, y) := \|x - y\|_p$, and $\|x\|_p^p := \sum_i |x(i)|^p$ for $p < \infty$, while $\|x\|_\infty := \sup_i |x(i)|$.

A continuous analog of this metric is provided by the strongly related example

$$X = C[a\,..\,b], \quad d(f, g) := \|f - g\|_p$$

with

$$\|f\|_p := \Big(\int_a^b |f(t)|^p \,\mathrm{d}t\Big)^{1/p}$$

for $1 \le p \le \infty$. Again, the case $p = \infty$ is meant as a limit,

$$\|f\|_\infty := \sup_{a \le t \le b} |f(t)|.$$

There are analogous formulæ when the interval $[a\,..\,b]$ is replaced by some suitable subset of $\mathbb{R}^n$.

In all of these examples, you should *verify* (e.g., by looking up the argument) that the proposed metric is, in fact, a metric. In this, it is usually hardest to verify the triangle inequality. Correspondingly, it is the most useful property of a metric.

Note that *any subset of a ms is itself a ms.* E.g., $(\mathbb{Z}, |\cdot|)$ is an instructive *example* of a ms.

A metric space (X, d) is a ts, with the nbhdsystem for $x \in X$ given by

$$\mathbf{B}(x) := \{B_r(x) : r > 0\},$$

or by the equivalent, but simpler

$$\mathbf{B}(x) := \{B_{1/m}(x) : m = 1, 2, \ldots\}.$$

Here,

$$B_r(x) := \{y \in X : d(y,x) < r\}$$

is the **open ball of radius** r **and center** x.

Think of $d(x,y)$ as the **distance** between the points or elements x and y. Then

$$d(x,Y) := \inf_{y\in Y} d(x,y)$$

gives the distance of x from the subset Y of X. Note the use of "inf" instead of "min" here. This implies that $\exists\{(y_n) \text{ in } Y\}$ s.t. $\lim_{n\to\infty} d(x,y_n) = d(x,Y)$. But this by itself does not guarantee that there is some $y \in Y$ for which $d(x,y) = d(x,Y)$. If such a y exists, we call it a **best approximation** (=: **ba**) **to** x **from** Y. See (III.13) for an example of a lss that is closed yet fails to provide a b.a. to any x (not in that lss).

We define

$$B_r(Y) := \{x \in X : d(x,Y) < r\}, \qquad B_r^-(Y) := \{x \in X : d(x,Y) \le r\}$$

the **open**, resp. **closed ball of radius** r **and center** Y. Note that $Y_1 \subset Y_2$ implies $B_r(Y_1) \subset B_r(Y_2)$. This implies that $B_r(Y) \cup B_r(Z) \subset B_r(Y \cup Z)$, while, conversely, for any $x \in B_r(Y \cup Z)$, we must have $r' := d(x, Y \cup Z) < r$, hence, for all $s > r'$, there must be $v \in Y \cup Z$ with $d(x,v) < s$, and, in particular, there must be v in either Y or Z so that $d(x,v) < r$, i.e., x must be in $B_r(Y) \cup B_r(Z)$. This proves that $B_r(Y \cup Z) = B_r(Y) \cup B_r(Z)$, and, in particular,

$$B_r(Y) = \bigcup_{y\in Y} B_r(y).$$

H.P.(9) Verify that $B_r(Y)$ deserves the epithet "open", and $B_r^-(Y)$ the epithet "closed" (Hint: Show that $d(\cdot, Y)$ is continuous), but that $B_r^-(Y) \ne B_r(Y)^-$ is possible. (Hint: X = integers with the absolute value metric).

H.P.(10) Verify that $\mathbf{B}$ as defined for a ms satisfies the two properties (2) and (5) of a neighborhood system. (Hint, consider $Y = \{t\}$ in the previous homework.)

The closure Y^- of a subset Y of X can be characterized in terms of d:

$$\begin{aligned} x \in Y^- &\iff \forall\{\varepsilon > 0\}\ \exists\{y \in Y\}\ \ d(x,y) < \varepsilon \\ &\iff d(x,Y) = 0 \\ &\iff \forall\{r > 0\}\ x \in B_r(Y) \\ &\iff x \in \bigcap_{r>0} B_r(Y). \end{aligned} \tag{10}$$

Thus, the last term shows an example of an intersection of open sets that is not open.

Y is **bounded** $:= \exists\{r, x\}\ Y \subseteq B_r(x)$. Since $B_r(x) \subseteq B_{r+d(x,y)}(y)$, the boundedness does not depend on the choice of x, but the radius r certainly will. We could measure the size of Y by the smallest possible r (even if it does not exist). Instead, one measures the **size** of such a set Y by the **diameter** of Y:

$$\operatorname{diam} Y := \sup_{x,y\in Y} d(x,y) \le 2\inf\{r : Y \subseteq B_r(x)\}.$$

E.g., $\operatorname{diam} B_r(x) \le 2r$.

H.P.(11) Prove that the last two inequalities could be strict. (Hint: $(\mathbb{Z}, |\cdot|)$.)

(11) Lemma. $\operatorname{diam} Y = \operatorname{diam} Y^-$.

Proof. $\forall\{x, y \in Y^-,\ r > 0\}\ \exists\{x', y' \in Y\}\ \ d(x, x'), d(y, y') < r$, therefore

$$|d(x,y) - d(x',y')| \le |d(x,y) - d(x,y')| + |d(x,y') - d(x',y')| \le 2r.$$

Conclusion: $\{d(x,y) : x, y \in Y^-\} \subseteq \{d(x',y') : x', y' \in Y\}^-$. Therefore

$$\operatorname{diam} Y^- \le \sup_{x',y' \in Y} d(x',y') = \operatorname{diam} Y \le \operatorname{diam} Y^-,$$

the first inequality using the fact that $\sup Z = \sup Z^-$ for $Z \subseteq \mathbb{R}$ (cf. (8)). □

H.P.(12) Prove that $\forall\{z,\ Y \subseteq X\}\ \ Y \subseteq B^-_{d(z,Y)+\operatorname{diam} Y}(z)$.

H.P.(13) Show that (11)Lemma is a special case of (8)Proposition.

**** modulus of continuity ****

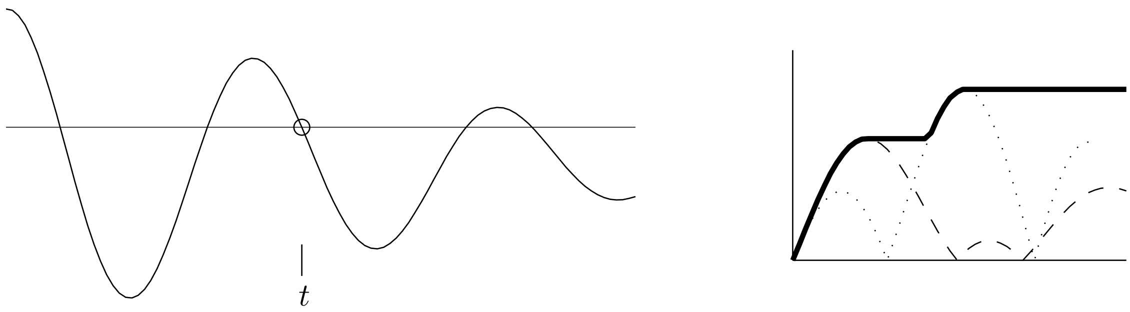
In a ms, we have the following **quantitative** description of continuity. For T, U ms's, $t \in T$, and $f : T \to U$, the function

$$\omega_{f,t} : \mathbb{R}_+ \to \mathbb{R}_+ : h \mapsto \sup_{d(s,t)<h} d(f(s), f(t))$$

is the **modulus of continuity** of f **at** t. If the supremum is also taken over t, we get

$$\omega_f(h) := \sup_{t \in T} \omega_{f,t}(h) = \sup_{d(s,t)<h} d(f(s), f(t)),$$

the **(uniform) modulus of continuity** of f. These moduli are nondecreasing, nonnegative functions, hence $\omega_{f,t}(0+) := \lim_{h\to 0} \omega_{f,t}(h)$ and $\omega_f(0+)$ are well-defined.



(12) Figure. Construction of $\omega_{f,t}$ (heavy) as the 'sunrise function' $h \mapsto \max\{f_r(s), f_l(s) : s \le h\}$ of $f_r : h \mapsto |f(t+h) - f(t)|$ (dashed) and $f_l : h \mapsto |f(t-h) - f(t)|$ (dotted).

f is continuous at t iff $\omega_{f,t}(0+) = 0$; f is **uniformly** continuous iff $\omega_f(0+) = 0$. The modulus of continuity gives the continuity information of practical interest: If we want to ensure that the uncertainty in $f(N)$ is $< r$, i.e., that $f(N) \subseteq B_r(f(t))$, then we must choose N in $B_w(t)$ with $w \le \omega_{f,t}^{-1}(r)$. (Here we set $\omega^{-1}(r) := \infty$ in case $\sup_h \omega(h) < r$.)

H.P.(14) Sketch $\omega_{f,0}$ on $[0\,..\,1]$ for f that vanishes at 0 and agrees with (i) $t \mapsto \sin(1/t)$ (ii) $t \mapsto t\sin(1/t)$ for $t \neq 0$. Also, sketch ω_f on $[0\,..\,3]$ for $f := (1+()^2)^{-1}$.

H.P.(15) For $f : [a\,..\,b] \to \mathbb{R}$ and $t := (a = t_1 < t_2 < \cdots < t_{\ell+1} = b)$, let $P_t f$ denote the broken line interpolant to f at t, i.e., $P_t f$ is the unique continuous piecewise linear function on $[a\,..\,b]$ with breakpoints $t_2, \ldots, t_\ell$ that agrees with f at t. Prove that $\sup_{x\in[a..b]} |f(x) - P_t f(x)| \le \omega_f(\max_j (t_{j+1} - t_j))$.

The moduli of continuity of two functions are close if the functions are uniformly close. Precisely:

(13) Lemma. $\omega_{f,t} \;\le\; \omega_{g,t} + 2d_\infty(f,g)$, *with*

$$d_\infty(f,g) := \sup_{t\in T} d(f(t), g(t)).$$

Proof:
$$\begin{aligned} d(f(s), f(t)) &\le d(f(s), g(s)) + d(g(s), g(t)) + d(g(t), f(t)) \\ &\le d_\infty(f,g) + d(g(s), g(t)) + d_\infty(f,g). \end{aligned}$$
□

(14) Corollary. *If $f : T \to U$ is in the closure, in the sense of the metric d_∞, of $C(T \to U)$, then $f \in C(T \to U)$.*

Proof: By the lemma, $\omega_{f,t} \le \omega_{g,t} + 2d_\infty(f,g)$, and therefore $\omega_{f,t}(0+) \le 2d_\infty(f,g)$ for every continuous g. Since, by assumption, I can make $d_\infty(f,g)$ as small as I please, $\omega_{f,t}(0+) = 0$ follows. □

Continuous functions are classified by just how fast $\omega_f(h)$ approaches 0 as $h \to 0$. For example, we say that f is **Lipschitz** continuous in case $\omega_f(h) \le \kappa h$, all h, for some constant κ, called a **Lipschitz constant** for f.

E.g., the function $d(\cdot, Y)$ is Lipschitz continuous (with constant 1). Any continuously differentiable function f on the interval $[a\,..\,b]$ is Lipschitz continuous since

$$|f(s) - f(t)| = |\int_t^s Df(u)\,\mathrm{d}u| \le \|Df\|_\infty |s-t|,$$

with the Lipschitz constant equal to the absolute max of the derivative. For $0 < \alpha < 1$, the function $()^\alpha : [0\,..\,1] \to [0\,..\,1] : t \mapsto t^\alpha$ fails to be Lipschitz continuous since its modulus of continuity is $\omega_{()^\alpha}(h) = h^\alpha$.

H.P.(16) Prove that $f : [0\,..\,1]^n \to \mathbb{R}$ is constant in case $\omega_f \le \mathrm{const}()^\alpha$ for some $\alpha > 1$.

H.P.(17) Prove that $d(\cdot, Y)$ is Lipschitz continuous with constant 1.

** convergence of sequences **

To be precise, a **sequence** $(x_n) = (x_1, x_2, \ldots)$ in a ms X is a map

$$x : \mathbb{N} \to X : n \mapsto x_n$$

from the natural numbers into X.

We say that (x_n) **converges to** y and write $\lim x_n = y$ in case $y \in X$ and $\lim d(x_n, y) = 0$. If also $\lim x_n = z$, then $d(y,z) \le d(y,x_n) + d(x_n,z) \xrightarrow[n\to\infty]{} 0$, so $y = z$. If $\operatorname{ran} x \subseteq Y$ (i.e., $\forall\{n\}\ x_n \in Y$) and $y = \lim x_n$, then $y \in Y^-$. In other words, convergence of sequences in a metric space behaves just like convergence of sequences of real numbers.

Remarks. It is worthwhile (in view of more general topologies you might run into later) to pursue a little further the (rigorous) view that a sequence (x_n) in the ms X is the map $x : \mathbb{N} \to X : n \mapsto x_n$, and so to think of convergence as a question of "continuity at infinity". For, we have $y = \lim x_n$ exactly when

$$\forall\{\varepsilon > 0\}\, \exists\{n \in \mathbb{N}\}\, \forall\{m > n\} \quad x_m \in B_\varepsilon(y).$$

Hence, if we define

$$>n := \{m \in \mathbb{N} : m > n\},$$

then this says that

$$\forall\{\varepsilon > 0\}\, \exists\{n \in \mathbb{N}\} \quad x(>n) \subseteq B_\varepsilon(y),$$

with

$$x(>n) = x_{>n} = \{x_m : m > n\}$$

the n**-tail** of the sequence x. So, even more succinctly, $\lim x_n = y$ iff

$$x\mathbf{B}(\infty) \succ \mathbf{B}(y),$$

where

$$\mathbf{B}(\infty) := \{>n : n \in \mathbb{N}\}$$

plays the role of the neighborhood system "at ∞" (even though $\infty \notin \mathbb{N}$).

This makes explicit that, as far as the limit is concerned, the essential feature of a sequence x is the collection of its n-tails.

In these terms, it is obvious that, for $f : X \to U$ continuous at $y = \lim x_n$, we have $\lim f(x_n) = f(y)$; it's just the continuity at ∞ of the composite map $\mathbb{N} \xrightarrow{x} X \xrightarrow{f} U$. Of course, you can (and should) verify this directly.

Note that *a convergent sequence is necessarily* **bounded**, i.e., lies entirely in some ball, since, for every $r > 0$, some n-tail must lie in $B_r(x_\infty)$, hence the entire sequence must lie in the ball of radius $r + \max_{j \le n} d(x_j, x_\infty)$ around the limit, x_∞.

Convergence of a sequence makes sense in a general ts X: We say that $x : \mathbb{N} \to X$ **converges to** y in case $\{x_{>n} : n \in \mathbb{N}\} \succ \mathbf{B}(y)$. But ms's (and certain other ts's) are special in that the topology can be characterized in terms of sequence behavior. For example, the closure Y^- of a subset Y in the ms X is its **sequential closure**, i.e., the set of all limits of sequences in Y. Here is another example:

(15) Lemma. *X, Y ms's, $f : X \to Y$. Then*

$$f \text{ is continuous at } z \quad \Longleftrightarrow \quad \lim_n x_n = z \text{ implies } \lim_n f(x_n) = f(z).$$

Proof: We just proved '$\Longrightarrow$'. The proof of '$\Longleftarrow$' is by contradiction: If f fails to be continuous at z, then $\exists\{r > 0\}\ \forall\{s > 0\}\ f(B_s(z)) \not\subseteq B_r(f(z))$, i.e., $\exists\{r > 0\}\ \forall\{s > 0\}\ \exists\{x_s \in B_s(z)\}$ with $f(x_s) \notin B_r(f(x))$, i.e., $\lim_{s\to 0} x_s = z$ yet $f(x_s) \notin B_r(f(z))$ for any s. □

In more general ts's, for example in $C[a \mathinner{..} b]$ with the topology $\mathbf{B}_{pw}$ (see (4)Example) of pointwise convergence, the convergence behavior of sequences is not sufficient to characterize the topology (see H.P.(19)). More general 'sequences' called 'nets', or, equivalently, 'filter bases' are needed. The latter concept seems more straightforward: A **filter basis** is a collection $\mathbf{C}$ of nonempty subsets that satisfies (2)Neighborhood Assumption 1, i.e., $\mathbf{C} \succ \mathbf{C} \Cap \mathbf{C}$. The filter basis **$\mathbf{C}$ converges to t** if $\mathbf{C} \succ \mathbf{B}(t)$. This is related to the fact that, in ts's more general than ms's, it may be impossible to find an equivalent nbhdsystem that is totally ordered by inclusion.

H.P.(18) Prove that convergence of a sequence $n \mapsto f_n$ in $(\mathbb{F}^T, \mathbf{B}_{pw})$ is, indeed, **pointwise convergence**, i.e., $f = \lim_{n\to\infty} f_n$ in this topology if and only if, for all $t \in T$, the numerical sequence $n \mapsto f_n(t)$ converges to the number $f(t)$.

H.P.(19) Consider the space $X := b(T)$ of all bounded real-valued functions on some *uncountable* set T (e.g., $T = [0 \mathinner{..} 1]$), with the topology of pointwise convergence (cf. (4)Example and preceding homework). Let $X_0 := \{f \in X : \#\operatorname{supp} f < \infty\}$. Prove that X_0 is dense in X, but that the sequential closure of X_0 is a proper subset of X, namely $X_c := \{f \in X : \operatorname{supp} f \text{ is countable}\}$.

Application: Contraction maps and fixed point iteration

(16) Problem. *X ms, $f : X \to Y$ continuous. Given $y \in Y$, find $x \in X$* s.t. *$f(x) = y$.*

One can usually solve this basic problem by turning the given equation $f(?) = y$ into an equivalent **fixed point** equation

$$? = g(?)$$

for some continuous $g : X \to X$ (which depends on the given y, of course), and then generating a sequence (x_n) of approximations to the solution x by **fixed point iteration**,

$$x_{n+1} := g(x_n), \qquad n = 0, 1, 2, \ldots,$$

starting with some **initial guess** x_0.

Example 1. $X = \mathbb{R}^m, f = A \in \mathbb{R}^{m\times m}$, i.e., we are solving a linear system of equations. This can be brought into equivalent fixed point form by a **splitting** $A = M - N$, with M invertible. Then $Mx = Nx + y$, or, $x = M^{-1}Nx + M^{-1}y =: g(x)$. I trust that your Numerical Analysis background includes a discussion of this fixed point iteration.

Example 2. $X = C[a \mathinner{..} b], f(x) := x - \int_a^b k(\cdot, t)x(t)\,\mathrm{d}t$, leads to an **integral equation of the second kind**. If the **kernel** k is suitable, then fixed point iteration with $g(x) := \int_a^b k(\cdot, t)x(t)\,\mathrm{d}t + y$ will converge to the solution of the integral equation. I come back to this later, in Chapter VIII.

Example 3. The equation $x^2 = a$ (with $x, a \in \mathbb{R}_+$) has the well known equivalent fixed point form

$$x = g(x) := (x + a/x)/2.$$

Starting at any positive x_0, the resulting fixed point iterants $x_0, x_1, x_2, \ldots$ seem to cluster at the unique fixed point of g, i.e., at the unique x with $x = g(x)$.

Is there such a point?

If the sequence (x_n) generated by fixed point iteration converges, to some x_∞, then, by the assumed continuity of g, $g(x_\infty) = x_\infty$. So, in general, the limit of such an iteration sequence is a sought-for fixed point. But how can we ensure convergence?

A computationally effective way is to prove (if possible) that g is a (proper) **contraction**, i.e., g is Lipschitz continuous with Lipschitz constant $\kappa < 1$. This means that

$$\forall\{x, z \in X\}\ d(g(x), g(z)) \le \kappa d(x, z)$$

for some $\kappa < 1$. In these circumstances, *g has at most one fixed point*, i.e., we have **uniqueness**: Indeed, if both $x = g(x)$ and $z = g(z)$, then

$$d(x, z) = d(g(x), g(z)) \le \kappa d(x, z) \le d(x, z),$$

and this shows that $\kappa d(x, z) = d(x, z)$ which, given that $|\kappa| < 1$, is possible only if $d(x, z) = 0$, i.e., $x = z$.

Further,

$$d(x_{n+1}, x_n) = d(g(x_n), g(x_{n-1})) \le \kappa d(x_n, x_{n-1}),$$

and so, by induction,

$$d(x_{n+1}, x_n) \le \kappa^r d(x_{n+1-r}, x_{n-r}), \qquad r = 1, 2, \ldots.$$

So, in particular,

$$d(x_{n+1}, x_n) \le \kappa^n d(x_1, x_0).$$

Consequently, for $m > n$,

$$\begin{aligned} d(x_m, x_n) \ &\le \ d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \ \cdots \ + d(x_{n+1}, x_n) \\ &\le \big(\quad \kappa^{m-1} \quad + \quad \kappa^{m-2} \quad + \cdots + \quad \kappa^n \quad \big) d(x_1, x_0) \end{aligned}$$

i.e.

$$d(x_m, x_n) \ \le \ \kappa^n \frac{\kappa^{m-n} - 1}{\kappa - 1} d(x_1, x_0). \tag{17}$$

This shows (given that $0 \le \kappa < 1$) that

$$\limsup_{m,n\to\infty} d(x_m, x_n) = 0,$$

or, equivalently, that

$$\lim_{n\to\infty} \text{diam}\ \ x_{>n} = 0.$$

A sequence having this property is said to be a **Cauchy sequence**, or, for short, to be **Cauchy**. If we know nothing else, then we cannot conclude anything else.

For *example*, the set $X := \mathbb{Q} \cap [1\,..\,2]$ of rational numbers in the interval $[1\,..\,2]$ is a metric space under the usual metric $d(x, y) := |x - y|$. Starting with $x_0 = 2$, fixed point iteration with the function

$$g(x) := (x + 2/x)/2$$

of Example 3 (with $a = 2$) generates a sequence (x_n) of numbers all in X. This sequence is easily seen to be Cauchy since $|Dg| \le 1/2$ (on $[1\,..\,2]$), therefore $|g(x) - g(y)| \le |x - y|/2$ (there). Yet the sequence fails to converge, since its limit would have to be the number $\sqrt{2}$ and this number is not rational.

This motivates the discussion of

**** completeness ****

For any sequence $x : \mathbb{N} \to X$ in a ms X, diam $x_{>n}$ decreases (or, at least, does not increase) as n increases. Therefore, $\lim_n$ diam $x_{>n} = 0$ iff $\forall\{r > 0\}\ \exists\{n\}$ s.t. diam $x_{>n} < r$, i.e., s.t. , for all $m, m' > n$, $d(x_m, x_{m'}) < r$.

If $x : \mathbb{N} \to X$ converges, say to y, then $\forall\{r > 0\}\ \exists\{n\}$ s.t. $x_{>n} \subseteq B_r(y)$, therefore, x is Cauchy. In a complete ms, the converse holds as well.

(18) Definition. *X ms is* **complete** *:= every Cauchy sequence converges (to some point in X).*

The failing of the particular metric space $\mathbb{Q} \cap [1\,..\,2]$ in the earlier example, or of $\mathbb{Q}$ itself, is that it is not complete. By completing the rationals, we obtain the reals. This process of completing a metric space consists of adjoining, if necessary, points to the space to provide a limit for every Cauchy sequence. This can always be done in a reasonable way, i.e., so that the metric can be extended in a reasonable way to this larger set. Here is the naive approach.

We start off with the collection X' of all Cauchy sequences in X. We can embed X in X' by associating $y \in X$ with the constant sequence $\mathbb{N} \to X : n \mapsto y$. On X', we define

$$d' : X' \times X' : (x, y) \mapsto \lim_{n\to\infty} d(x_n, y_n),$$

and this is well-defined since

$$d(x_n, y_n) - d(y_m, x_m) = d(x_n, y_n) - d(x_m, y_n) \ + \ d(x_m, y_n) - d(x_m, y_m),$$

hence $\operatorname{diam}\{d(x_m, y_m) : m > n\} \le \operatorname{diam} x_{>n} + \operatorname{diam} y_{>n}$, showing that the *real* sequence $n \mapsto d(x_n, y_n)$ is Cauchy, hence converges. On the other hand, while d' is symmetric and satisfies the triangle inequality, it fails to be definite since, for any two sequences x and y with the same limit, $d'(x, y) = 0$.

To take care of this, one groups the elements of X' into classes:

$$\langle x\rangle := \{y \in X' : d'(y, x) = 0\},$$

and observes that these classes are disjoint (since $y \in \langle x\rangle \implies \langle y\rangle \subseteq \langle x\rangle$), hence the map

$$\lim : X' \to X'' := \{\langle x\rangle : x \in X'\} : x \mapsto \langle x\rangle$$

is well-defined.

It is now easy to verify (using the continuity of d, i.e., the fact that $x \to x'$ and $y \to y'$ implies that $d(x, y) \to d(x', y')$) that

$$d'' : X'' \times X'' : (\langle x\rangle, \langle y\rangle) \mapsto d'(x, y)$$

is well-defined (in particular, does not depend on the particular representatives $x \in \langle x\rangle$ and $y \in \langle y\rangle$ used here in $d'(x, y)$), and a metric, and reduces to d on X as embedded in X'', and X is dense in X'', and, finally and most importantly, X'' is complete. This says that, whatever new Cauchy sequences we might now be able to build in addition to those already in X, their limit is already in X'', i.e., we need not go through this process again.

H.P.(20) Prove that the map $X \to X'' : y \mapsto \langle(y)\rangle$ is onto if and only if X is complete.

H.P.(21) Given the Cauchy sequence $n \mapsto \langle x^{(n)}\rangle$ in X'', determine its limit. (Hint: Show first that one may assume wlog that the diameters of the tailends of the sequence $x^{(n)}$ go to zero uniformly in n; then try the sequence $(x_n^{(n)} : n \in \mathbb{N})$.)

H.P.(22) Prove: *The ms X is complete iff every decreasing sequence $A_1 \supseteq A_2 \supseteq \cdots$ of nonempty closed sets with* $\operatorname{diam} A_n \to 0$ *has a nontrivial intersection.*

From a practical point of view, this only provides a gain in neatness since these additional points can usually not be constructed anyway, but must be approximated by the terms of a sequence converging to it, i.e., by the terms of a Cauchy sequence that gave rise to that point in the first place. Don't underestimate neatness, though. Completeness will lead us to the uniform boundedness principle which allows us to connect convergence to stability in various Numerical Analysis processes.

The scalars $\mathbb{F}$ with the usual metric are complete. More generally, $\mathbb{F}^m$ with any of the p-metrics mentioned earlier is complete. More generally:

(19) Proposition. *The metric space*

$$b(T)$$

of all bounded real (or complex) functions on T with the **uniform metric**

$$d_\infty(f,g) := \|f-g\|_\infty$$

is complete.

Proof: If $x : \mathbb{N} \to b(T)$ is Cauchy, then, for all $t \in T$, $\mathbb{N} \to \mathbb{F} : n \mapsto x_n(t)$ is scalar Cauchy, hence, by the completeness of $\mathbb{F}$, the function

$$x_\infty : T \to \mathbb{F} : t \mapsto \lim x_n(t)$$

is well defined. Further, for any t,

$$(x_\infty - x_n)(t) = (x_\infty - x_m)(t) + (x_m - x_n)(t),$$

therefore, for any $m \ge n$,

$$(x_\infty - x_n)(t) \le |(x_\infty - x_m)(t)| + \operatorname{diam}\ x_{\ge n}.$$

By letting $m \to \infty$, this shows that

$$\|x_\infty - x_n\|_\infty \le \operatorname{diam}\ x_{\ge n} \quad \xrightarrow[n\to\infty]{} \quad 0.$$

In particular, $\|x_\infty\|_\infty \le \|x_n\|_\infty + \operatorname{diam}\ x_{\ge n} < \infty$, i.e., $x_\infty \in b(T)$, and so $x_\infty = \lim x_n$ (in the metric of $b(T)$). □

(20) Lemma. *A closed subset of a complete ms is complete.*

In particular, *if T is ms, then $bC(T) := b(T) \cap C(T)$ is complete*, since *$bC(T)$ is a closed subset of $b(T)$*, by the Corollary to (13)Lemma.

**** summary on contraction ****

(21) Proposition. *A proper contraction g (with Lipschitz constant κ) on a complete metric space X has exactly one fixed point, y say. This point is the limit of any sequence $x_0, x_1, x_2, \ldots$ of iterates generated by fixed point iteration starting from an arbitrary initial guess in X. The distance of x_n from y can be estimated by*

$$d(y, x_n) \le d(x_1, x_0)\kappa^n/(1-\kappa).$$

This estimate follows from (17) by letting $m \to \infty$.

You see here a good opportunity for using the *equivalence of topologies*, specifically the equivalence of certain metrics. For, while equivalent metrics give the same convergent sequences, the Lipschitz constant κ depends very much on the details of the particular metric. Here are two classical examples.

Linear iteration If $g : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto Mx$ for some $M \in \mathbb{R}^{n\times n}$, then its Lipschitz constant is the norm $\|M\| := \sup_x \|Mx\|/\|x\|$ of the matrix M with respect to the norm $\|\cdot\|$ that provides the metric for $\mathbb{R}^n$ (see the next chapter, particularly the material on approximate inverses). The central result concerning such **linear iteration** is that it converges iff $\|M\| < 1$ in *some* norm.

Picard iteration occurs in the discussion of **first-order ODE**s: Assume $f : [0\,..\,b]\times[-M\,..\,M] \to \mathbb{R}$ continuous, and uniformly Lipschitz continuous in its second argument, i.e., $\exists\{\kappa\}\ \forall\{r \in [0\,..\,b], s, t \in [-M\,..\,M]\}\ |f(r,s) - f(r,t)| < \kappa|s-t|$. Given the initial value c, we are seeking

$$x \in C^{(1)}[0\,..\,b] \text{ s.t. } x(0) = c,\ Dx(t) = f(t, x(t)) \text{ for } t \in [0\,..\,b]. \tag{22}$$

Picard's iteration function

$$g(x) := c + \int_0^{\cdot} f(s, x(s))\,\mathrm{d}s$$

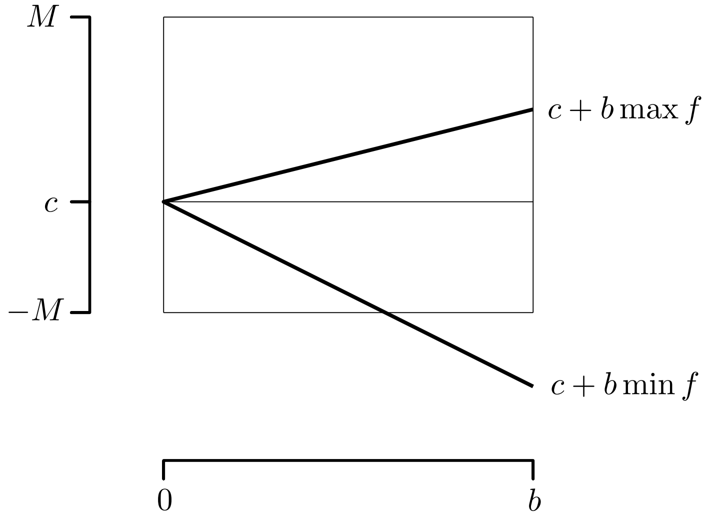
seems a natural one for the construction of a solution since x satisfies (22) exactly when x is a fixed point of g. But, in the simple metric d_∞, we only get

$$d_\infty(g(x), g(y)) \ \le\ \max_t \int_0^t |f(s,x(s)) - f(s,y(s))|\,\mathrm{d}s \le b\kappa d_\infty(x,y),$$

hence get a proper contraction only for sufficiently small κ and/or b. Now consider the *equivalent* metric $d(x,y) := \max_t \mathrm{e}^{-at}|x(t) - y(t)|$. Since

$$\int_0^t |f(s,x(s)) - f(s,y(s))|\,\mathrm{d}s \le \int_0^t \mathrm{e}^{as}\kappa\mathrm{e}^{-as}|x(s) - y(s)|\,\mathrm{d}s \le \int_0^t \mathrm{e}^{as}\,\mathrm{d}s\ \kappa d(x,y),$$

we obtain $d(g(x), g(y)) \le \max_t \mathrm{e}^{-at}\kappa(\mathrm{e}^{at} - 1)/a\ d(x, y)$. Hence, with a sufficiently large (e.g., $a = 2\kappa$), g is a proper contraction, and so (21)Proposition becomes applicable, providing for existence and uniqueness of a solution of (22), provided we can ensure that $x([0\,..\,b]) \subset [-M\,..\,M]$ implies that also $g(x)([0\,..\,b]) \subset [-M\,..\,M]$. For this (cf. (23)Figure), observe that $g(x)(s) - c \in s[\min f\,..\,\max f]$, with, e.g., $\min f := \min f([0\,..\,b] \times [-M\,..\,M])$, hence we are ok in case $c + b[\min f\,..\,\max f] \subset [-M\,..\,M]$. Otherwise, we either have to cut down b and/or try a more sophisticated analysis (using a Gronwall inequality).



(23) Figure. Bounds relevant in Picard iteration.

H.P.(23) Prove that the two metrics used in the Picard example lead to equivalent topologies.

Compactness and total boundedness

**** limit points ****

In the best of circumstances, a numerical method generates a Cauchy sequence, i.e., a sequence (x_n) with $\mathrm{diam}(x_{>n}) \to 0$. At times, though, one has to be satisfied with a sequence (x_n) that settles down in the much weaker sense that the set $\bigcap_{n\in\mathbb{N}} \left(x_{>n}\right)^-$ of its **limit point**s is not empty. If $y = \lim x_n$, then $d(y, x_{>n}) = 0$ for all n, therefore $y \in \bigcap_{n\in\mathbb{N}} \left(x_{>n}\right)^-$. In fact, in this case (as for any Cauchy sequence), $\mathrm{diam}(x_{>n})^- = \mathrm{diam}(x_{>n}) \to 0$, hence y is the only limit point of (x_n). In general, a sequence may have fewer or more limit points.

For *example*, the sequence given by $x_n := (-1)^n$, all n, has both 1 and -1 as its limit points, and none other, while the sequence given by

$$x_n := n/10^k \qquad \text{if } 10^{k-1} \le n < 10^k$$

has every point in $[(.1)\,..\,1]$ as limit point (and none other), and the sequence given by $x_n = n$, all n, has none.

H.P.(24) Prove that *any limit point of a sequence is a closure point of its range*, but that the converse does not hold.

Equivalently, y is a limit point of the sequence (x_n) iff $d(y, x_{>n}) = 0$ for all n. Thus, y is a limit point iff, for every n, there is an $m_n > m_{n-1}$ (with $m_0 := 0$) so that $d(y, x_{m_n}) < 1/n$ (since, whatever m_{n-1} might be, the m_{n-1}-tail of x would have y in its closure). In more conventional terms, this states that some **subsequence** of (x_n) converges to y. We call the ms X **sequentially compact** if every sequence has limit points.

A *standard example* of a sequence (x_n) without limit points is one for which $r := \inf_{n \neq m} d(x_n, x_m) > 0$, since no ball of radius $< r/2$ can contain more than one point from that sequence.

H.P.(25) Prove: (i) *Any Cauchy sequence having a limit point converges*; and (ii) *A sequence converges iff every subsequence has the same nonempty set of limit points.*

For any *real* sequence $n \mapsto x_n$, the related sequence $n \mapsto \sup x_{>n}$ is monotone nonincreasing, hence converges if it is bounded below. In that case, its limit is called the **limit superior** of (x_n), in symbols

$$\limsup_{n\to\infty} x_n := \lim_{n\to\infty} \sup x_{>n}.$$

Correspondingly, the **limit inferior** is the limit of the sequence $n \mapsto \inf x_{>n}$. Any limit point of (x_n) lies between these the two extreme limit points of the sequence.

**** compactness ****

Completeness ensures existence of a limit in case the sequence is Cauchy. Sequential compactness ensures the existence of a limit point for every sequence. I now discuss the relationship of sequential compactness of a ms to the more general notion of compactness of a subset of a ts. There are various equivalent definitions of compactness. The standard definition is: The subset Y of the ts $(X, \mathbf{B})$ is **compact** := *Every open cover* $\mathbf{O}$ *for* Y *contains a finite subcover.* This phrase means that, if

$$Y \subseteq \cup\mathbf{O} := \bigcup_{O\in\mathbf{O}} O$$

for some collection $\mathbf{O}$ of open sets, then already

$$Y \subseteq \cup\mathbf{O}'$$

for some *finite* subcollection $\mathbf{O}'$ of $\mathbf{O}$.

This is a technically convenient definition. For example, it is easy (as you should *verify*) to prove the

(24) Lemma. *If $f \in C(X \to U)$ and $Y \subseteq X$ is compact, then $f(Y)$ is compact.*

or the

(25) Lemma. *Any closed subset Y of a compact set X is compact.*

Proof: Any open cover for the closed set Y becomes an open cover for the entire set after adjunction of the open set $X\backslash Y$. After removal, if need be, of $X\backslash Y$ from the resulting finite subcover for the compact set X, we must have a finite cover for Y, from the original cover for Y. □

Here is a final example of the efficiency of the above definition of compactness.

(26) Lemma. *A continuous function on a compact ms is uniformly continuous.*

Proof: Let $f \in C(T \to U)$, with T, U ms's, and T compact. We must show that $\omega_f(0+) = 0$. Let $\varepsilon > 0$ be arbitrary. We must prove that $\omega_f(h) < \varepsilon$ for some $h > 0$ (hence $\omega_f(h') < \varepsilon$ for all $h' \le h$). Since f is continuous, all the numbers $h_t := (\omega_{f,t})^{-1}(\varepsilon/2)$ are positive, hence $\{B_{h_t/2}(t) : t \in T\}$ is an open cover for T. Since T is compact, there must already be a finite W so that $\{B_{h_t/2}(t) : t \in W\}$ covers T. It follows that $h := \min_{t\in W} h_t/2$ is positive, and that, for any $d(s,t) < h$, there is some $w \in W$ with $t \in B_{h_w/2}(w)$, therefore both t and s are in $B_{h_w}(w)$, hence $d(f(s), f(t)) \le d(f(s), f(w)) + d(f(w), f(t)) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. □

On the other hand, the purpose to which compactness is most often put is easier to discern if you write down the equivalent definition in terms of closed sets, recalling that a set is closed iff its complement is open. This equivalent definition of the compactness of Y reads: *If Y fails to have any point in common with $\cap\mathbf{A}$* (for some collection $\mathbf{A}$ of closed sets), *then there must already be a* finite *subcollection $\mathbf{A}' \subseteq \mathbf{A}$ so that $\cap\mathbf{A}'$ has no point in common with Y.* To put this implication the other way: *In order to know that $(\cap\mathbf{A}) \cap Y \ne \{\}$ for some collection $\mathbf{A}$ of closed sets, it is sufficient to know that $\mathbf{A}$ is* **centralized wrto** Y, *i.e., has the socalled*

(27) Finite Intersection property w.r.to Y. $\forall\{$ *finite* $\mathbf{A}' \subseteq \mathbf{A}\}\ (\cap\mathbf{A}') \cap Y \ne \{\}$.

This *equivalent characterization of the compactness of* Y shows that compactness is associated with the *existence* of certain points: Think of each $A \in \mathbf{A}$ as the set of points satisfying a certain condition. Then compactness allows us to conclude the existence of a point in Y satisfying *infinitely* many conditions as soon as we can show that *every* choice of *finitely* many conditions is satisfied by some point in Y. For example,

(28) Proposition. *If $f \in C(X \to \mathbb{R})$ and X is compact, then there is $x \in X$ s.t. $f(x) = \sup f(X)$.*

Proof. By definition of $M := \sup f(X) := \sup_{x\in X} f(x)$, we have

$$\forall\{r < M\}\ \exists\{x \in X\}\ f(x) \ge r.$$

This says that the collection of sets

$$f^{-1}[r\,..\,M], \quad r < M,$$

has the Finite Intersection Property, while the continuity of f ensures that each such set is closed (as the inverse image of a closed set). Therefore, by compactness, one can find $x \in \cap_{r<M} f^{-1}[r\,..\,M]$, i.e., $\exists\{x \in X\}\ \forall\{r < M\}\ r \le f(x) \le M$. But this says that $f(x) = M$. □

Remark. The full force of continuity of f is not needed here. It is only required that $f^{-1}[r\,..\,\infty)$ be closed for every r. A real-valued function having this property is called **upper semicontinuous** (why?). By considering $-f$ instead, we also find $x \in X$ for which $f(x) = \inf f(X)$, and can do that already if f is merely **lower semicontinuous**, i.e., $f^{-1}(-\infty\,..\,r]$ is closed for every r.

(29) Corollary. *For a compact ms T, $bC(T) = C(T)$.*

Here are further examples of the use of compactness for proving existence:

(30) Proposition. *A compact set is sequentially compact.*

Indeed, for any sequence $x : \mathbb{N} \to X$, the collection of closed sets $(x_{>n})^-$ is decreasing (i.e., $(x_{>n})^- \supseteq (x_{>m})^-$ for $n \le m$), hence trivially has the Finite Intersection Property, therefore $\bigcap_{n\in\mathbb{N}}(x_{>n})^-$ is nonempty by compactness, i.e., the sequence has limit points.

H.P.(26) Prove that, in a ms, any limit point of a sequence is the limit of some subsequence (of that sequence). Can you give an example of a ts in which this conclusion does not hold? (see H.P. (V.15))

(31) Proposition. *A compact set Y in a ms X provides best approximations for every $x \in X$.*

Indeed, apply (28)Proposition to the function $f : Y \to \mathbb{R} : y \mapsto -d(x, y)$. In fact, this conclusion can already be reached if Y is merely **boundedly compact**, i.e., $B_r^-(x) \cap Y$ is compact for every (finite) r and x.

We observed earlier (see (25)Lemma) that a closed subset of a compact set is compact. The complementary statement holds under additional assumptions.

(32) Lemma. *If $Y \subseteq X$ is compact and X is, e.g., a ms, then Y is closed.*

Proof: Let $z \in Y^-$. Then every nbhd of z intersects Y, hence so does the closure of every nbhd of z, therefore, by (2),

$$\mathbf{B}^-(z) := \{N^- : N \in \mathbf{B}(z)\}$$

has the FIP wrto Y. Since Y is compact, this implies that $(\cap\mathbf{B}^-(z)) \cap Y \ne \{\}$. Hence, any assumption (e.g., that X is a ms) that guarantees that $\cap\mathbf{B}^-(z) = \{z\}$ finishes the proof. □

(33) Note. A ts with the property that $\cap\mathbf{B}^-(x) = \{x\}$, or, equivalently, the property that $y \ne z \implies \{\} \in \mathbf{B}(y)\Cap\mathbf{B}(z)$, is called **Hausdorff**. In particular, *any metric space is Hausdorff*, since $y \ne z$ implies that $2r := d(y, z) > 0$, hence $B_r(y)$ and $B_r(z)$ are nonintersecting nbhds of y and z, respectively.

H.P.(27) (a) Prove: *The set $\{x\}$ in the ts X is closed iff $\cap\mathbf{B}(x) = \{x\}$.* A ts in which $\{x\}$ is closed for every $x \in X$ is called a T_1**-space**.

(b) Show by an example that a T_1-space need not be a T_2**-space**, i.e., Hausdorff.

(c) Show by an example that a T_0**-space**, i.e., a space in which, for every $x \ne y$, some nhbd of x excludes y or else some nbhd of y excludes x, need not be a T_1-space.

H.P.(28) Prove: *The ts X is Hausdorff iff every filter basis has at most one limit.* (Hint: If $\{\} \notin \mathbf{B}(x)\Cap\mathbf{B}(y)$, then the collection of all finite intersections of nbds of x and of y is a filter basis that converges to both x and y.)

Finally, *if Y is closed and compact, then Y is compact in any Z containing it*, since then, for any subset A of Z closed in Z, $A \cap Y$ is closed.

H.P.(29) Prove: *X, U ms's, X compact. $f : X \to U$ 1-1, onto, continuous $\implies$ f^{-1} continuous.*

** total boundedness **

In a ms (X, d) (and in certain other ts's), compactness is closely related to total boundedness. Recall that a subset $Y \subseteq X$ is bounded if $Y \subseteq B_r(x)$ for some r and some x.

(34) Definition. *$Y \subseteq X$ is* **totally bounded** $:= \forall\{r > 0\}\ \exists\{\text{finite } U \subseteq X\}$ s.t. $Y \subseteq B_r(U) := \bigcup_{x\in U} B_r(x)$. *Any such finite set U for given r is called an r-***net** *for Y.*

Any convergent sequence (or, more precisely, the set consisting of the entries of a convergent sequence) is totally bounded. In contrast, any set containing a sequence (x_n) with $\inf_{n\ne m} d(x_n, x_m) > 0$ fails to be totally bounded. For example, the elements $e_n : \mathbb{N} \to \mathbb{R} : i \mapsto \delta_{in}$ of the ms ℓ_1 are each at distance 2 from each other, yet all lie in the bounded set $B_1^-(0)$, hence the latter fails to be totally bounded.

H.P.(30) Prove: *If Y is totally bounded, then the required r-nets can always be chosen from Y.* (But, in verifying total boundedness, it is at times handy not to have to insist on that.)

(35) Theorem. *(X, d) ms. X is compact iff X is complete and totally bounded.*

Proof: compact $\implies$ complete: Any sequence (x_n) in X has a limit point, say y, by compactness. If the sequence is also Cauchy, then $d(x_{n+1}, y) \le \operatorname{diam}(x_{>n})^- = \operatorname{diam} x_{>n} \xrightarrow[n\to\infty]{} 0$.

compact $\implies$ totally bounded: $\forall\{r > 0\}$ the collection $\{B_r(x) : x \in X\}$ is an open cover for X, hence must contain a finite subcover, by compactness.

complete & totally bounded $\implies$ compact: The proof is an abstract version of the familiar argument by contradiction for the Heine-Borel Theorem, as follows.

Suppose that $\mathbf{O}$ is an open cover of X containing no finite subcover. Now, by assumption, any set in X can be covered by *finitely* many balls of a given radius, hence if each of these balls were finitely coverable, then so would be the entire set. It follows that any $Z \subseteq X$ that cannot be finitely covered from $\mathbf{O}$ contains subsets of arbitrarily small (positive) diameter that also cannot be finitely covered from $\mathbf{O}$. This permits inductive construction of a decreasing sequence $X = Z_1 \supseteq Z_2 \supseteq \cdots$ of sets with $\lim_n \operatorname{diam}(Z_n) = 0$ and none finitely coverable from $\mathbf{O}$; in particular, none is empty. However, X is complete by assumption, hence (see H.P. (22)) there is some $z \in \cap_n Z_n^-$, and therefore, certainly, some $O \in \mathbf{O}$ containing z and, O being open, therefore also containing $B_r(z)$ for some $r > 0$, hence $z \in Z_n^- \subset O$ as soon as $\operatorname{diam}(Z_n) < r$, showing that all but finitely many of the Z_n are finitely coverable from $\mathbf{O}$ after all. □

H.P.(31) Prove: *A totally bounded sequence (i.e., a sequence with a totally bounded range) has a Cauchy subsequence.* (You might have fun supplying the details for the following auxiliary argument: If there are no Cauchy subsequences, then $X := \{x_n : n \in \mathbb{N}\} = \operatorname{ran}(x_n)$ is complete, hence compact, hence (x_n) has convergent subsequences.)

(36) Corollary. *Let X ms. Then: X is compact $\iff$ X is sequentially compact.*

Proof. By (30)Prop., compact $\implies$ sequentially compact. For the converse, assume that X is not compact. Then X is not complete or else X is not totally bounded. If X is not complete, then it contains a Cauchy sequence without a limit, hence without a limit point (since, by H.P.(II.25), any limit point of a Cauchy sequence is necessarily its limit), so X is not sequentially compact. If X is not totally bounded, then, for some $r > 0$, no matter how we choose the finite set U, X fails to be covered by $B_r(U)$. This permits choice of a sequence $u_1, u_2, \ldots$ in X in that order so that

$$u_n \in X \backslash B_r(u_{<n}), \quad n = 1, 2, \ldots.$$

This implies that $d(u_n, u_m) \geq r$ for all $n \neq m$, hence (u_n) has no limit point. □

**** examples of compact sets ****

(37). *$X = \mathbb{R}^m$ with uniform metric, $Y \subseteq X$. Y is compact $\Longleftrightarrow$ Y is closed and bounded.*

Proof: '$\Longrightarrow$': true in any metric space.

'$\Longleftarrow$': X complete and Y closed $\Longrightarrow$ Y complete. Also, any bounded set in $\mathbb{R}^m$ is totally bounded. Here is a formal proof: Let

$$f : \mathbb{R}^m \to \mathbb{Z}^m : x \mapsto \big(\lfloor x(i) \rfloor\big)_{i=1}^m$$

(with $\mathbb{Z}$ the integers and $\lfloor t \rfloor :=$ the largest integer no bigger than t). Then $d(x, f(x)) \leq 1$ and $\# f(B_r(0)) < \infty$ for any finite r. Therefore, for any $r > 0$, $f_r : x \mapsto r f(x/r)$ satisfies $d(x, f_r(x)) \leq r$ and $f_r(B_s(0))$ is finite for any s. If now Y is bounded, then $Y \subseteq B_s(0)$ for some $s > 0$. But then, for any $r > 0$, $Y \subseteq B_r(U)$ with $U := f_{r/2}(B_s(0))$. □

Actually, this is a special case of

(38) Arzela-Ascoli. *$Y \subseteq X = C(T)$ with T a compact ms. Then Y compact $\Longleftrightarrow$ Y closed, bounded and* **equicontinuous**, *i.e., $\omega_Y := \sup_{f\in Y} \omega_f$ satisfies $\omega(0+) = 0$.*

Proof: '$\Longrightarrow$': Since Y is bounded, $Y \subseteq B_M(0)$ for some M, hence $\forall\{f \in Y\}\ \omega_f \leq 2M$. Therefore $\omega_Y := \sup_{f\in Y} \omega_f$ is well-defined and nondecreasing. Hence, $\omega_f(0+) = 0$ iff $\forall\{r > 0\}\ \exists\{h > 0\}\ \omega_Y(h) < r$.

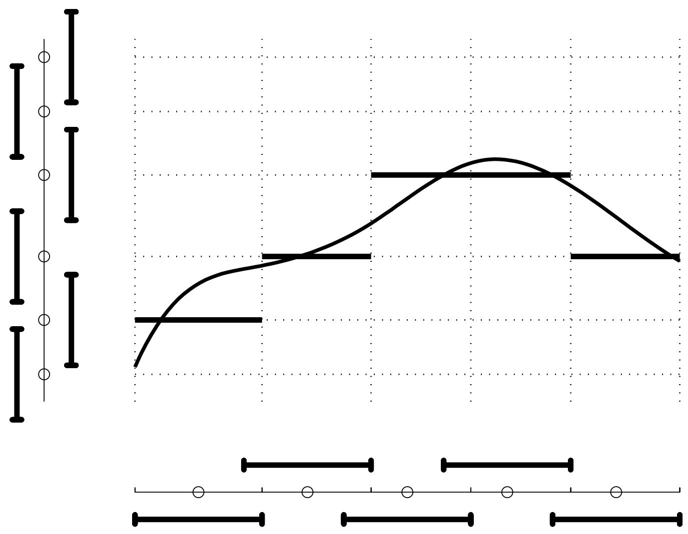
Let $r > 0$. For any $f \in C(T)$, $\omega_f(0+) = 0$ by the compactness of T (see (26)Lemma), hence there is $h_f > 0$ so that $\omega_f(h_f) < r$. Further, Y is totally bounded, hence there is some *finite* U so that $Y \subset B_r(U)$, while (by (13)Lemma) $\omega_{B_r(f)} \leq \omega_f + 2r$. Therefore, $h := \min_{f\in U} h_f > 0$ and

$$\omega_Y(h) \leq \max_{f\in U} \omega_{B_r(f)}(h) \leq \max_{f\in U} \omega_f(h) + 2r \leq 3r,$$

and that is sufficient.

To be sure, the total boundedness is essential here. For example, for $T = [0\,..\,1]$ and $Y = \{()^\alpha : \alpha > 0\}$, we have $\omega(h) = \sup_{\alpha>0} h^\alpha = 1$, hence $\omega(0+) = 1 \neq 0$ in this example.

'$\Longleftarrow$': Since the metric of $C(T)$ is the restriction to $C(T)$ of the metric of $b(T)$, it is sufficient to prove that Y is a compact subset of $b(T)$. For this, Y is complete (as a closed subset of the complete ms $C(T)$) by (20)Lemma, and is totally bounded, by the following Lemma.



(39) Figure. An r-net for a totally bounded set of functions.

(40) Lemma. *If T is totally bounded, then any bounded and equicontinuous subset Y of $b(T)$ is totally bounded.*

Proof: Construct an r-net for Y in $b(T)$ as follows. Let $h := \omega^{-1}(r/2)$ and let $\{t_1, \ldots, t_s\}$ be an h-net for T. Further, let

$$E_i := B_h(t_i) \backslash B_h(t_{<i}), \qquad i = 1, \ldots, s.$$

Since Y is bounded, there is some finite interval I containing $Y(T)$. Choose an $r/2$-net W for I. We claim that the (finite!) collection U of functions with values in W and constant on each E_i is an r-net for Y: If $f \in Y$, let g be the function in U whose value on E_i is within $r/2$ of $f(t_i)$, all i. Then, for $t \in E_i$, $d(t, t_i) < h$, therefore

$$|g(t) - f(t)| \le |g(E_i) - f(t_i)| + |f(t_i) - f(t)| < r/2 + r/2.$$

□

Remark. The essence of the proof is that the cartesian product of totally bounded sets is totally bounded (in the **product metric** $d((p,q),(p',q')) := d(p,p') + d(q,q')$).

H.P.(32) Extend the above proof of (40)Lemma to the case where $b(T)$ denotes all bounded functions on the totally bounded ms T into some **boundedly compact** ms R, i.e., a complete ms in which closed and bounded sets are totally bounded (instead of into $\mathbb{R}$).

** Tykhonov **

In full generality, the topology of pointwise convergence (recall (4)Example) is not metrizable, i.e., is not equivalent to a topology derived from some metric. This means that we must look for different ways to establish compactness in such spaces. The basic result is

(41) Tykhonov's Theorem. *The cartesian product $X := \times_{t\in T} X_t$ of an arbitrary (nonempty) indexed collection $(X_t : t \in T)$ of compact ts's is compact in the product topology.*

To be sure, we can think of $x \in X$ as a map on T that associates $t \in T$ with some $x(t) \in X_t$, all t. Further, the **product topology** is, in effect, that of pointwise convergence. Precisely, for $x \in X$,

$$\mathbf{B}(x) := \{\underset{t\in T}{\times} N_t : \ N_t \begin{cases} \in \mathbf{B}(x(t)), & t \in S, \\ = X_t & \text{otherwise} \end{cases}\Bigg\}, S \subseteq T, \#S < \infty\}.$$

Proof: Let $\mathbf{A}$ be a nonempty collection of closed sets having the FIP (= Finite Intersection Property). We show that $\bigcap \mathbf{A} \neq \{\}$.

For this, let $\mathbf{C}$ be a largest collection of sets having the FIP and containing $\mathbf{A}$. (Such a maximal collection exists by the Hausdorff Maximality Theorem; see the latter's use in the proof of the (IV.27) Hahn-Banach Theorem.) Then, $\mathbf{C}$ contains any set containing some set in $\mathbf{C}$. More interestingly, $\mathbf{C}$ is closed under finite intersections, i.e., contains the intersection of any finitely many of its elements. Most interestingly, $\mathbf{C}$ contains any $C \subseteq X$ that intersects every set in $\mathbf{C}$: indeed, the adjoining of such a set to $\mathbf{C}$ will not destroy the FIP since, for any finite subset $\mathbf{C}'$ of $\mathbf{C}$, also $\bigcap \mathbf{C}'$ is in $\mathbf{C}$, hence is intersected by C.

Now set, for $t \in T$ and $C \subseteq X$,

$$C|_t := \{x(t) : x \in C\}.$$

Then, for each $t \in T$, $\{(C|_t)^- : C \in \mathbf{C}\}$ is a collection of closed sets having the FIP, hence, by the compactness of X_t, there is some $a(t) \in \bigcap\{(C|_t)^- : C \in \mathbf{C}\}$. This implies that every neighborhood N_t of $a(t)$ intersects $C|_t$ for every $C \in \mathbf{C}$, therefore, the set

$$\underset{s\in T}{\times} \begin{Bmatrix} N_t & s = t, \\ X_s & s \neq t \end{Bmatrix}$$

intersects every $C \in \mathbf{C}$, hence is in $\mathbf{C}$, by the latter's maximality, hence so is the intersection of finitely many such sets. Since the latter are the neighborhoods of a, this implies that every neighborhood of a intersects every $C \in \mathbf{C}$, hence $a \in \bigcap_{C\in\mathbf{C}} C^-$. In particular, $a \in \bigcap \mathbf{A}$. □

We will only need here the following special case.

(42) Proposition. *For an arbitrary (nonempty) set T and arbitrary closed and bounded intervals I_t, the set $X := \times_{t\in T} I_t$ is compact in the topology of pointwise convergence.*

