Chapter 7. Some Applications

Contraction Maps and Picard's Theorem

7.1 Definition: Let X be a metric space. A map $f : X \to X$ is called a **contraction map** on X when there exists a constant $c \in [0,1)$ such that for all $x,y \in X$ we have

$$d\big(f(x),f(y)\big) \leq c\,d(x,y).$$

Such a constant c is called a **contraction constant** for f. Note that every contraction map is uniformly continuous.

7.2 Definition: For a map $f : X \to X$ (where X is any set), a point $a \in X$ such that $f(a) = a$ is called a **fixed point** of f.

7.3 Theorem: *(The Banach Fixed-Point Theorem) Every contraction map on a complete metric space has a unique fixed point.*

Proof: Let X be a complete metric space and let $f : X \to X$ be a contraction map on X with contraction constant $c \in [0,1)$. Let $x_0 \in X$ be any point. Let $x_1 = f(x_0)$ and $x_2 = f(x_1) = f^2(x_0)$ and so on, so that for $n \geq 1$ we have $x_n = f(x_{n-1}) = f^n(x_0)$. Note that the sequence $(x_n)_{n\geq 0}$ is Cauchy because for $n < m$ we have

$$\begin{aligned} d(x_n,x_m) &= d\big(f^n(x_0),f^n(x_{m-n})\big) \leq c^n\,d\big(x_0,x_{m-n}\big) \\ &\leq c^n\big(d(x_0,x_1)+d(x_1,x_2)+\cdots+d(x_{m-n-1},x_{m-n})\big) \\ &\leq c^n d(x_0,x_1)\big(1+c+c^2+\cdots+c^{m-n-1}\big) \\ &\leq c^n d(x_0,x_1)\,\tfrac{1}{1-c} \;\longrightarrow 0 \text{ as } n\to\infty. \end{aligned}$$

Since X is complete, the sequence $(x_n)_{n\geq 0}$ converges, so we can let $a = \lim\limits_{n\to\infty} x_n$. Note that $f(a) = a$ since f is continuous at a so $f(a) = f\big(\lim\limits_{n\to\infty} x_n\big) = \lim\limits_{n\to\infty} f(x_n) = \lim\limits_{n\to\infty} x_{n-1} = a$. Finally note that for $a,b \in X$, if $f(a) = a$ and $f(b) = b$ then since

$$d(a,b) = d\big(f(a),f(b)\big) \leq c\,d(a,b)$$

with $0 \leq c < 1$, it follows that $d(a,b) = 0$ so that $a = b$.

7.4 Example: Define $f : [2,\infty) \to [2,\infty)$ by $f(x) = x + \frac{1}{x}$. Note that $f'(x) = 1 - \frac{1}{x^2}$ so that $\frac{3}{4} \leq f'(x) < 1$ for all $x \in [2,\infty)$. By the Mean Value Theorem, given $x,y \in [2,\infty)$ we can choose c between x and y such that $f(x)-f(y) = f'(c)(x-y)$, and then we have $\big|f(x)-f(y)\big| = \big|f'(c)\big|\,|x-y| < |x-y|$. Thus f has the property that $\big|f(x)-f(y)\big| < |x-y|$ for all $x,y \in [2,\infty)$, but it is not a contraction map, and f has no fixed point because $f(x) = x + \frac{1}{x} > x$ for all $x \in [2,\infty)$.

7.5 Example: Define $f : \big[0,\frac{\pi}{3}\big] \to \big[0,\frac{\pi}{3}\big]$ by $f(x) = \cos x$ $\Big($note that $\cos(0) = 1$ and $\cos\big(\frac{\pi}{3}\big) = \frac{1}{2}$ and $\cos x$ is decreasing, so we have $f\big(\big[0,\frac{\pi}{3}\big) = \big[\frac{1}{2},1\big] \subseteq \big[0,\frac{\pi}{3}\big]\Big)$. Since $\big|f'(x)\big| = \sin x$ which is increasing on $\big[0,\frac{\pi}{3}\big]$, we have $0 \leq \big|f'(x)\big| \leq \frac{\sqrt{3}}{2}$ for all $x \in \big[0,\frac{\pi}{3}\big]$. By the Mean Value Theorem (as above) we have $\big|f(x)-f(y)\big| \leq \frac{\sqrt{3}}{2}|x-y|$ for all $x,y \in \big[0,\frac{\pi}{3}\big]$ so that f is a contraction map with contraction constant $c = \frac{\sqrt{3}}{2}$. By the Banach Fixed-Point Theorem, f has a unique point, that is there is a unique $a \in \big[0,\frac{\pi}{3}\big]$ such that $\cos a = a$. The proof of the theorem shows that we can find a as follows: choose any $x_0 \in \big[0,\frac{\pi}{3}\big]$ and let $x_n = f(x_{n-1}) = \cos(x_{n-1})$ for $n \geq 1$, and then $x_n \to a$.

7.6 Definition: Let $A \subseteq \mathbb{R}^2$ and let $f : A \to \mathbb{R}$. We say that f satisfies a **Lipschitz condition** on A when there exists a constant $\ell \geq 0$ such that for all $x, y_1, y_2 \in \mathbb{R}$ for which $(x, y_1) \in A$ and $(x, y_2) \in A$, we have

$$\big|f(x, y_2) - f(x, y_1)\big| \leq \ell\,|y_2 - y_1|.$$

Such a constant ℓ is called a **Lipschitz constant** for f.

7.7 Theorem: *(Picard) Let U be an open set in $\mathbb{R}^2$, let $(a, b) \in U$, and let $F : U \to \mathbb{R}$ satisfy a Lipschitz condition on U. Then there exists $\delta > 0$ such that the differential equation $\frac{dy}{dx} = F(x, y)$ has a unique solution $y = f(x)$ with $f(a) = b$, defined for all $x \in [a-\delta, a+\delta]$.*

Proof: First note that $y = f(x)$ is a solution to the differential equation $\frac{dy}{dx} = F(x, y)$ with $f(a) = b$ if and only if $f(x)$ satisfies the integral equation

$$f(x) = b + \int_a^x F\big(t, f(t)\big)\,dt$$

for all $x \in [a-\delta, a+\delta]$. Let ℓ be a Lipschitz constant for F. Choose $r > 0$ such that $\overline{B}\big((a, b), r\big) \subseteq U$ and let $k = \max\limits_{(x,y)\in\overline{B}((a,b),r)} \big|F(x, y)\big|$. Choose δ with $0 < \delta < \frac{1}{\ell}$ small enough such that the rectangle $R = [a-\delta, a+\delta] \times [b-k\delta, b+k\delta]$ is contained in $B\big((a, b), r\big)$. Verify as an exercise (Using the Mean Value Theorem) that if $f(x)$ is any solution to the given differential equation with $f(a) = b$ then the graph of f must be contained in the rectangle R. Let

$$X = \big\{f \in \mathcal{C}[a-\delta, a+\delta] \,\big|\, \text{Graph}(f) \subseteq R\big\}.$$

Verify that X is a closed subspace of the metric space $\mathcal{C}[a-\delta, a+\delta]$ (using the supremum metric) and so X is complete. Define $G : X \to \mathcal{C}[a-\delta, a+\delta]$ by

$$G(f)(x) = b + \int_a^x F\big(t, f(t)\big)\,dt.$$

Note that $G(X) \subseteq X$ because for all $f \in X$ and $x \in [a-\delta, a+\delta]$ we have

$$\big|G(f)(x) - b\big| = \left|\int_a^x F(t, f(t))dt\right| \leq \left|\int_a^x k\,dt\right| = k|x - a| \leq k\delta.$$

Note that G is a contraction map on X, with contraction constant $c = \ell\delta < 1$ because, for all $f, g \in X$ and all $x \in [a-\delta, a+\delta]$, we have

$$\begin{aligned}
\big|G(f)(x) - G(g)(x)\big| &= \left|\int_a^x \big(F(t, f(t)) - F(t, g(t))\big)dt\right| \leq \left|\int_a^x \big|F(t, f(t)) - F(t, g(t))\big|dt\right| \\
&\leq \left|\int_a^x \ell\big|f(t) - g(t)\big|\,dt\right| \leq \left|\int_a^x \ell\,\|f - g\|_\infty\,dt\right| \\
&= \ell|x - a|\,\|f - g\|_\infty \leq \ell\delta\,||f - g||_\infty.
\end{aligned}$$

By the Banach Fixed-Point Theorem, the map G has a unique fixed point $f \in X$, and this function $f \in X$ is the unique solution to the above integral equation, which is equivalent to the given differential equation.

7.8 Theorem: *(Peano) Let $U \in \mathbb{R}^2$ be open, let $(a, b) \in U$, and let $F : U \to \mathbb{R}$ be continuous. Then there exists $d > 0$ such that the differential equation $\frac{dy}{dx} = F(x, y)$ has a solution $y = f(x)$ which is defined for all $x \in [a-d, a+d]$.*

Proof: The proof is a bit too hard for this course, but it is included in Appendix 1.

7.9 Definition: Let X be a metric space. For $a \in X$ and $\emptyset \neq B \subseteq X$, we define the **distance** between a and B to be

$$d(a,B) = \inf\left\{d(a,x)\,\middle|\,x \in B\right\} = \inf\left\{d(x,a)\,\middle|\,x\in B\right\} = d(B,a).$$

7.10 Theorem: *Let X be a metric space, let $a \in X$, and let $\emptyset \neq B \subseteq X$.*

(1) The function $g : X \to \mathbb{R}$ given by $g(x) = d(a,x)$ is uniformly continuous.
(2) If B is compact then there exists $b \in B$ such that $d(a,B) = d(a,b)$.

Proof: Note that for $x, y \in X$, by the Triangle Inequality we have $g(x) = d(a,x) \leq d(a,y) + d(y,x) = g(y) + d(x,y)$ so that $g(x) - g(y) \leq d(x,y)$. Similarly (by interchanging x and y) we have $g(y) - g(x) \leq d(x,y)$ so that $d\big(g(x), g(y)\big) = \big|g(x) - g(y)\big| \leq d(x,y)$. It follows that g is uniformly continuous: indeed given $\epsilon > 0$ we can choose $\delta = \epsilon$ and then for all $x, y \in X$ with $d(x,y) < \delta$, we have $d\big(fgx), g(y)\big) \leq d(x,y) < \delta = \epsilon$. This proves Part (1).

To prove Part (2), suppose that B is compact. Note that since $g : X \to \mathbb{R}$ is continuous and $B \subseteq X$, the restriction $g : B \subseteq X \to \mathbb{R}$ is also continuous. Since B is compact and $g : B \to \mathbb{R}$ is continuous, it follows from the Extreme Value Theorem that g attains its minimum, so we can choose $b \in B$ such that $g(b) = \min\left\{g(x)\,\middle|\,x\in B\right\}$, that is

$$d(a,b) = \min\left\{d(a,x)\,\middle|\,x\in B\right\} = d(a,B).$$

7.11 Definition: Let X be a metric space. For two nonempty subsets $\emptyset \neq A, B \subseteq X$, we define the **Hausdorff distance** between A and B to be

$$D(A,B) = \max\Big\{\sup\left\{d(a,B)\middle|\,a\in A\right\}, \sup\left\{d(A,b)\middle|\,b\in B\right\}\Big\}.$$

Let $\mathcal{K} = \mathcal{K}(X)$ be the set of nonempty compact subsets of X. Note that when $A, B \in \mathcal{K}$, by the previous theorem we have

$$D(A,B) = \max\Big\{\max\left\{d(a,B)\middle|\,a\in A\right\}, \max\left\{d(A,b)\middle|\,b\in B\right\}\Big\}.$$

7.12 Theorem: *Let X be a normed linear space. The Hausdorff distance D is a metric on the set $\mathcal{K} = \mathcal{K}(X)$ of nonempty compact subsets of X. This metric is called the* **Hausdorff metric** *on $\mathcal{K}$.*

Proof: It is clear that D is symmetric (that is $D(A,B) = D(B,A)$ for all $A, B \in \mathcal{K}$. We claim that D is positive definite. It is clear that $D(A,B) \geq 0$ for all $A, B \in \mathcal{K}$, so we need to verify that if $D(A,B) = 0$ then $A = B$. Let $A, B \in \mathcal{K}$ and suppose that $D(A,B) = 0$, so that we have $\max_{a\in A} d(a,B) = 0 = \max_{b\in B} d(b,A)$. Since $\max_{a\in A} d(a,B) = 0$, we have $d(a,B) = 0$ for all $a \in A$, hence $a \in B$ for all $a \in A$, hence $A \subseteq B$. Similarly, since $\max_{b\in A} d(b,A) = 0$ we have $B \subseteq A$ and so $A = B$, as required.

It remains to prove that D satisfies the Triangle Inequality. To this end, we introduce some notation. For $A \in \mathcal{K}$ and $r \geq 0$, let

$$\begin{aligned} A_r &= \big\{x\in X \,\big|\, d(x,A) \leq r\big\} = \big\{x\in X \,\big|\, \exists\, a\in A\ d(a,x) \leq r\big\} \\ &= \big\{x\in X \,\big|\, \exists\, a\in A\ x\in \overline{B}(a,r)\big\} = \bigcup_{a\in A} \overline{B}(a,r). \end{aligned}$$

Note that A_r is closed because the function $g : X \to \mathbb{R}$ given by $g(x) = d(x,A)$ is continuous and we have $A_r = g^{-1}\big([0,r]\big)$.

Using this notation, it is clear that when $A, B \in \mathcal{K}$ and $r \geq 0$, if $A \subseteq B$ then $A_r \subseteq B_r$. We claim that when $A \in \mathcal{K}$ and $r, s \geq 0$ we have $(A_r)_s = A_{r+s}$. Let $x \in (A_r)_s$. Choose $y \in A_r$ such that $d(y,x) \leq s$. Since $y \in A_r$ we can choose $a \in A$ such that $d(a,y) \leq r$. Then $d(a,x) \leq d(a,y) + d(y,x) \leq r + s$ and so we have $x \in A_{r+s}$. This shows that $(A_r)_s \subseteq A_{r+s}$. Now let $x \in A_{r+s}$. Choose $a \in A$ such that $d(a,x) \leq r+s$. Let $[a,x]$ be the line segment in X from a to x. For any point $y \in [a,x]$ we have $d(a,y) + d(y,x) = d(a,x) \leq r + s$. Choose a point y on this line segment so that $d(a,y) \leq r$ and $d(y,x) \leq s$. Since $d(a,y) \leq r$ we have $y \in A_r$, and since $d(y,x) \leq s$ we have $x \in (A_r)_s$. This shows that $A_{r+s} \subseteq (A_r)_s$, so we have $(A_r)_s = A_{r+s}$ as claimed.

Now let us show that D satisfies the Triangle Inequality. Let $A, B, C \in \mathcal{K}$. Let $r = D(A,B)$ and $s = D(B,C)$. Since $D(A,B) = r$ we have $\max_{b\in B} d(b,A) \leq r$, so $d(b,A) \leq r$ for all $b \in B$, hence $b \in A_r$ for all $b \in B$, and so we have $B \subseteq A_r$. Similarly, since $D(B,C) = s$ we have $C \subseteq B_s$. Thus $C \subseteq B_s \subseteq (A_r)_s = A_{r+s}$. So for all $c \in C$ we have $d(c,A) \leq r+s$, and hence $\max_{c\in C} d(c,A) \leq r+s$. Similarly (by interchanging the roles of A and C) we have $\max_{a\in A} d(a,C) \leq r+s$, and so $D(A,C) \leq r+s$, as required.

7.13 Theorem: *When X is a complete normed linear space, $\mathcal{K}(X)$ is a complete metric space, using the Hausdorff metric. Indeed if $(A_n)_{n\geq 1}$ is a Cauchy sequence in $\mathcal{K}$ and if we let $C=\bigcap_{n=1}^\infty B_n$ where $B_n=\overline{\bigcup_{k=n}^\infty A_k}$, then we have $A_n\to C$ in $\mathcal{K}$.*

Proof: We use the notation from the previous proof: for $A\in\mathcal{K}$ and $r\geq 0$, we let

$$A_r=(A)_r=\left\{x\in X\,\middle|\,d(x,A)\leq r\right\}=\left\{x\in X\,\middle|\,\exists\,a\in A\ d(a,x)\leq r\right\}.$$

Note that A_r is compact: it is closed because the function $g(x)=d(x,A)$ is continuous and $A_r=g^{-1}[0,r]$, and it is bounded because A is bounded, and if we choose R so that $\|a\|\leq R$ for all $a\in A$ then, by the Triangle Inequality, we have $\|x\|\leq R+r$ for all $x\in A_r$.

Let $(A_n)_{n\geq 1}$ be a Cauchy sequence in $\mathcal{K}$. Since $(A_n)_{n\geq 1}$ is Cauchy, we can choose $m\in\mathbb{Z}^+$ so that $n,\ell\geq m\implies D(A_n,A_\ell)\leq 1$. For $n\geq m$ we have $D(A_n,A_m)\leq 1$ so that $A_n\subseteq(A_m)_1$. Since A_m is compact, it is bounded, and so $(A_m)_1$ is also bounded (if $\|x\|\leq R$ for all $x\in A_m$ then $\|y\|\leq R+1$ for all $y\in(A_m)_1$). Since the sets $A_1,\cdots,A_{m-1}$ are all bounded, and $A_n\subseteq(A_m)_1$ for all $n\geq m$, we can choose a closed ball $M\subseteq X$ such that $A_n\subseteq M$ for all $n\in\mathbb{Z}^+$.

For each $n\in\mathbb{Z}^+$, let $B_n=\overline{\bigcup_{k=n}^\infty A_k}$. Note that each B_n is compact with $M\supseteq B_1\supseteq B_2\supseteq B_3\supseteq\cdots$. Let $C=\bigcap_{n=1}^\infty B_n$. Note that $C\neq\emptyset$ since M has the finite intersection property on closed sets.

We claim that $B_n\to C$ in $\mathcal{K}$, or equivalently that $D(B_n,C)\to 0$ in $\mathbb{R}$. Since $B_1\supseteq B_2\supseteq\cdots$, the sequence $\big(D(B_n,C)\big)_{n\geq 1}$ is decreasing, so it converges. Let $r=\lim\limits_{n\to\infty}D(B_n,C)$. We have $r\geq 0$, and we need to show that $r=0$. For each $n\in\mathbb{Z}^+$, choose $b_n\in B_n$ so that $D(B_n,C)=d(b_n,C)$. Since M is compact, we can choose a convergent subsequence $(b_{n_k})_{k\geq 1}$ of $(b_n)_{n\geq 1}$, and let $c=\lim\limits_{k\to\infty}b_{n_k}$. Let $n\geq m$. Choose $\ell\in\mathbb{Z}^+$ so that $n_\ell\geq n$ and note that $B_{n_\ell}\subseteq B_n$. For all $k\geq\ell$ we have $b_{n_k}\in B_{n_k}\subseteq B_{n_\ell}\subseteq B_n$, and since B_n is closed, it follows that $c=\lim\limits_{k\to\infty}b_{n_k}\in B_n$. Since $c\in B_n$ for all $n\in\mathbb{Z}^+$, we have $c\in C$. Since $b_{n_k}\to c$ we have $D(B_{n_k},C)=d(b_{n_k},C)\leq d(b_{n_k},c)\to 0$, and so $r=\lim\limits_{k\to\infty}D(B_{n_k},C)=0$, as required.

We claim that $A_n\to C$. Let $\epsilon>0$. Since (A_n) is Cauchy, we can choose $m\in\mathbb{Z}^+$ such that $k,\ell\geq m\implies D(A_k,A_\ell)\leq\epsilon$. For all $k\geq m$, since $D(A_k,A_m)\leq\epsilon$ we have $d(x,A_m)\leq\epsilon$ for all $x\in A_k$ so that $A_k\subseteq(A_m)_\epsilon$. Since $A_k\subseteq(A_m)_\epsilon$ for all $k\geq m$ and $(A_m)_\epsilon$ is closed, we have

$$C=\bigcap_{n=1}^\infty B_n\subseteq B_m=\overline{\bigcup_{k=m}^\infty A_k}\subseteq\overline{(A_m)_\epsilon}=(A_m)_\epsilon.$$

Since $B_n\to C$ we can choose $\ell\in\mathbb{Z}^+$ such that $n\geq\ell\implies D(B_n,C)\leq\epsilon$. When $n\geq\ell$, since $D(B_n,C)\leq\epsilon$, we have $d(x,C)\leq\epsilon$ for all $x\in B_n$ so that $B_n\subseteq C_\epsilon$. For $n\geq\max\{m,\ell\}$. Since $C\subseteq(A_m)_\epsilon$, we have $d(c,A_m)\leq\epsilon$ for all $c\in C$ so that $\max_{c\in C}d(c,A_n)\leq\epsilon$, and since $A_n\subseteq B_n\subseteq C_\epsilon$, we have $d(a,C)\leq\epsilon$ for all $a\in A_n$ so that $\max_{a\in A_n}d(a,C)\leq\epsilon$. Thus for all $n\geq\max\{m,\ell\}$, $D(A_n,C)=\max\big\{\max\limits_{c\in C}d(c,A_n),\max\limits_{a\in A_n}d(a,C)\big\}\leq\epsilon$.

7.14 Definition: When W is a vector space, an **affine space** in W is a subset of the form $P = U + p = \{u + p \,|\, u \in U\}$ where $p \in W$ and U is a subspace of W. The **dimension** of the affine space $P = U + p$ is defined to be the dimension of the subspace U. When V and W are vector spaces, an **affine map** from V to W is a map $F : V \to W$ of the form $F(x) = L(x) + p$ where $p \in W$ and $L : V \to W$ is a linear map. In particular, an affine map from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a map $F : \mathbb{R}^n \to \mathbb{R}^m$ of the form $F(x) = Ax + b$ where $b \in \mathbb{R}^m$ and $A \in M_{m\times n}(\mathbb{R})$ (that is, A is an $m \times n$ matrix).

7.15 Remark: There are several different ways in which one can define the dimension of a subset of $\mathbb{R}^n$ in such a way that affine spaces in $\mathbb{R}^n$ have their appropriate dimension (which is a natural number), and a **fractal** can then be defined to be a subset of $\mathbb{R}^n$ which has a fractional dimension (that is a subset whose dimension is not a natural number).

7.16 Definition: An **isometry** (or a **distance-preserving map**) on $\mathbb{R}^n$ is defined to be a map $F : \mathbb{R}^n \to \mathbb{R}^n$ with the property that $\|F(x) - F(y)\| = \|x - y\|$ for all $x, y \in \mathbb{R}^n$. It can be shown (using linear algebra) that the isometries on $\mathbb{R}^n$ are the affine maps of the form $F(x) = Ax + b$ where $b \in \mathbb{R}^n$ and $A \in M_{n\times n}(\mathbb{R})$ with $A^TA = I$. A **similarity** (or a **distance-scaling map**) of **scaling factor** s on $\mathbb{R}^n$ is a map $F : \mathbb{R}^n \to \mathbb{R}^n$ with the property that $\|F(x) - F(y)\| = s\|x - y\|$ for all $x, y \in \mathbb{R}^n$. It can be shown (using linear algebra) that the similarities on $\mathbb{R}^n$ of scaling factor s are the affine maps of the form $F(x) = sAx + b$ where $b \in \mathbb{R}^m$ and $A \in M_{n\times n}(\mathbb{R})$ with $A^TA = I$.

7.17 Definition: A set $S \subseteq \mathbb{R}^n$ is **self-similar** when there exist similarities $F_1, F_2, \cdots, F_m$ on $\mathbb{R}^n$ with scaling factors $s_1, s_2, \cdots, s_m$ where $0 < s_k < 1$ and $m \geq 2$, such that

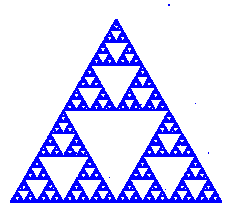
$$S = F_1(S) \cup F_2(S) \cup \cdots \cup F_m(S)$$

and the sets $F_k(S)$ are disjoint, except possibly for overlaps along simple boundaries, meaning that S is contained in a polyhedron P and the sets $F_k(S)$ only overlap along the boundaries of the polyhedra $F_k(P)$.

7.18 Example: The standard **Cantor set** $C \subseteq \mathbb{R}$ is obtained recursively as follows: Start with the interval $C_0 = [0, 1]$. Having obtained C_n, which is a union of 2^n disjoint closed intervals, each of length $\frac{1}{3^n}$, remove the open middle third of each closed interval to obtain C_{n+1}. Having constructed the sets $C_0, C_1, C_2, \cdots$, define $C = \bigcap_{n=0}^{\infty} C_n$. The Cantor set C is self-similar with $C = F_1(C) \cup F_2(C)$ where $F_1(x) = \frac{1}{3}x$ and $F_2(x) = \frac{1}{3}x + \frac{1}{3}$.

7.19 Example: The **Sierpinski Triangle** (also called the **Sierpinski gasket**) $T \subseteq \mathbb{R}^2$ is obtained recursively as follows: Start with the closed equilateral triangle T_0 with vertices at $(0, 0)$, $(1, 0)$ and $\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$, which has area $A_0 = \frac{\sqrt{3}}{8}$. Having obtained T_n, which is a union of 3^n closed equilateral triangles of area $A_n = \frac{1}{4^n}A_0$, and which only intersect at their vertices, divide each triangle into 4 congruent triangles and remove the interior of the middle one to obtain T_{n+1}. Having constructed the sets $T_0, T_1, T_2, \cdots$, define $T = \bigcap_{n=0}^{\infty} T_n$. The Sierpinsky triangle T is self-similar with $T = F_1(T) \cup F_2(T) \cup F_3(T)$ where

$$F_1\binom{x}{y} = \tfrac{1}{2}\binom{x}{y} \;,\; F_2\binom{x}{y} = \tfrac{1}{2}\binom{x}{y} \;,\; F_3\binom{x}{y} = \tfrac{1}{2}\binom{x}{y} + \tfrac{1}{4}\binom{1}{\sqrt{3}}.$$



7.20 Example: The **Sierpinski square** (also called the **Sierpinski carpet**) $S \subseteq \mathbb{R}^2$ is constructed like the Sierpinski triangle but beginning with the unit square, subdividing into 9 congruent squares, and removing the interior of the central square. The **Sierpinski cube** (also called the **Menger sponge**) $Q \subseteq \mathbb{R}^3$ is constructed by beginning with the unit cube, subdividing into 27 smaller cubes, then removing the centre cube of each face and the central cube, leaving the union of 20 cubes.

7.21 Example: The **Koch curve** $K \subseteq \mathbb{R}^2$ is constructed by beginning with the unit interval $[0,1]$ along the x-axis, subdividing into 3 subintervals of length $\frac{1}{3}$, constructing an equilateral triangle above the central subinterval, then removing that central subinterval. This gives a polygonal path which is the union of 4 intervals each of length $\frac{1}{3}$. We then repeat the procedure on each interval. The Koch curve K is self-similar with $K = F_1(K) \cup F_2(K) + F_3(K) + F_4(K)$ where

$$F_1\binom{x}{y} = \tfrac{1}{4}\binom{x}{y}\,,\ F_2\binom{x}{y} = \tfrac{1}{3}R_{\frac{\pi}{3}}\binom{x}{y} + \binom{\frac{1}{3}}{0}\,,\ F_3\binom{x}{y} = \tfrac{1}{3}R_{\frac{2\pi}{3}}\binom{x}{y} + \binom{1}{0}\,,\ F_4\binom{x}{y} = \tfrac{1}{3}\binom{x}{y} + \binom{\frac{2}{3}}{0}$$

where R_θ denotes the rotation given by $R_\theta\binom{x}{y} = \left(\begin{smallmatrix}\cos\theta & \sin\theta \\ \sin\theta & -\cos\theta\end{smallmatrix}\right)\binom{x}{y}$. The **Koch snowflake** is the union of 3 Koch curves, translated and rotated so they surround the three edges of an equilateral triangle. The Koch snowflake is not self-similar.

7.22 Note: The unit interval $I = [0,1] \subseteq \mathbb{R}$ is self-similar; it is the union of $m = n$ copies of itself scaled by $s = \frac{1}{n}$, and we say its dimension is $d = 1$. The unit square $I^2 \subseteq \mathbb{R}^2$ is self-similar; it is the union of $m = n^2$ copies of itself scaled by $s = \frac{1}{n}$, and we say its dimension is $d = 2$. The unit cube $I^3 \subseteq \mathbb{R}^3$ is self-similar; it is the union of $m = n^3$ copies of itself scaled by $s = \frac{1}{n}$, and we say its dimension is $d = 3$. Note that in each case, we have $m = \left(\frac{1}{s}\right)^d$ so that

$$d = \log_{1/s} m = \tfrac{\ln m}{\ln \frac{1}{s}} = -\tfrac{\ln m}{\ln s}.$$

More generally, if we partition $I = [0,1]$ into m subintervals of lengths $s_1, s_2, \cdots, s_m$ then we have $s_1 + s_2 + \cdots + s_m = 1$, and if we partition the unit square I^2 into m smaller squares with side-lengths $s_1, s_2, \cdots, s_m$ then we have ${s_1}^2 + {s_2}^2 + {s_3}^2 = 1$, and if we partition the unit cube I^3 into m smaller cubes with side lengths $s_1, s_2, \cdots, s_m$ then we have ${s_1}^3 + {s_2}^3 + \cdots + {s_m}^3 = 1$. In all cases we have

$${s_1}^d + {s_2}^d + \cdots + {s_m}^d = 1$$

for the appropriate dimension d. This motivates the following definition.

7.23 Definition: Let $S \subseteq \mathbb{R}^n$ be self-similar with $S = F_1(S) \cup S_2(S) \cup \cdots \cup F_m(S)$, where F_k is a similarity of scaling factor s_k, and where the sets $F_k(S)$ only overlap along simple boundaries. We define the **similarity dimension** of S to be the unique number $d = d(S) = d_{\text{sim}}(S)$ such that

$${s_1}^d + {s_2}^d + \cdots + {s_m}^d = 1.$$

In particular, in the case that $s_1 = s_2 = \cdots = s_m = s$ we have $m\,s^d = 1$ so that

$$d = \tfrac{\ln m}{\ln \frac{1}{s}} = -\tfrac{\ln m}{\ln s}.$$

7.24 Example: For the standard Cantor set C, we can take $m = 2$ and $s = \frac{1}{3}$ so that the similarity dimension is $d(C) = \frac{\ln 2}{\ln 3} \cong 0.631$.

For the Sierpinski triangle T, we can take $m = 3$ and $s = \frac{1}{2}$ so that $d(T) = \frac{\ln 3}{\ln 2} \cong 1.585$.

For the Sierpinski square S, we can take $m = 8$ and $s = \frac{1}{3}$ so that $d(S) = \frac{\ln 8}{\ln 3} \cong 1.893$.

For the Sierpinski cube Q, we can take $m = 20$ and $s = \frac{1}{3}$ so that $d(Q) = \frac{\ln 20}{\ln 3} \cong 2.727$.

For the Koch curve K, we can take $m = 4$ and $s = \frac{1}{3}$ so that $d(K) = \frac{\ln 4}{\ln 34} \cong 1.262$.

7.25 Remark: There are a number of alternate ways of defining dimension. For example, when $S \subseteq \mathbb{R}^n$ we can define the **capacity dimension** (also called the **box-counting dimension** or the **Hausdorff dimension**) of S to be

$$d = d_{\text{cap}}(S) = \lim_{r \to 0^+} \frac{\ln N_S(r)}{\ln \frac{1}{r}}$$

where $N_S(r)$ is the smallest number of n-boxes of side-length r whose union contains S. It can be shown (though the proof is not easy) that when $S \subseteq \mathbb{R}^n$ is self-similar we have $d_{\text{cap}}(S) = d_{\text{sim}}(S)$, so capacity dimension generalizes similarity dimension.

7.26 Exercise: Find the capacity dimension of the set $S = \left\{ \frac{1}{n} \,\middle|\, n \in \mathbb{Z}^+ \right\}$.

7.27 Note: When $S \subseteq \mathbb{R}^2$ is compact and self-similar or, more generally, when S is compact with $S = F_1(S) \cup F_2(S) \cup \cdots \cup F_m(S)$, where each F_k is a contraction map on $\mathbb{R}^2$, the following algorithm can be implemented on a computer to produce an accurate image of S: Let A_0 be an arbitrary finite set in $\mathbb{R}^2$ (for example $A_0 = \{0\}$) and, having constructed A_n, we let $A_{n+1} = F_1(A_n) \cup F_2(A_n) \cup \cdots \cup F_m(A_n)$. Then, as the following theorem shows, the sets A_n become arbitrarily close to the set S in the sense that for all $\epsilon > 0$ there exists $N \geq 0$ such that when $n \geq N$, every point in A_n is within a distance of ϵ to a point in S and vice versa.

7.28 Theorem: *Let $F_1, F_2, \cdots, F_\ell : \mathbb{R}^2 \to \mathbb{R}^2$ be contraction maps on $\mathbb{R}^2$. Let $\mathcal{K} = \mathcal{K}(\mathbb{R}^2)$.*
(1) The map $G : \mathcal{K} \to \mathcal{K}$ given by $G(A) = \bigcup_{k=1}^{\ell} F_k(A)$ is a contraction map on $\mathcal{K}$.
(2) There is a unique compact set S in $\mathbb{R}^2$ with $G(S) = S$. If A_0 is any nonempty compact set in $\mathbb{R}^2$ and we define $(A_n)_{n \geq 1}$ by $A_{n+1} = G(A_n)$, then $A_n \to S$ in $\mathcal{K}$.

Proof: First we claim that G is well-defined. Let $A \in \mathcal{K}$. Since A is compact and F_k is continuous (contraction maps are continuous), $F_k(A)$ is compact. Since each set $F_k(A)$ is closed and bounded, the set $G(A) = \bigcup_{k=1}^{\ell} F_k(A)$ is closed and bounded, and hence compact. This shows that when $A \in \mathcal{K}$ we also have $G(A) \in \mathcal{K}$, and so the map $G : \mathcal{K} \to \mathcal{K}$ is well-defined, as claimed.

For each k, let $c_k < 1$ be a contraction constant for F_k, and let $c = \max\{c_1, \cdots, c_\ell\}$. Note that $c < 1$. We claim that c is a contraction constant for G. Let $u \in G(A)$, say $u = F_k(a_k)$ where $a_k \in A$. Choose $b \in B$ so that $d(a, b) = d(a, B)$, and let $v = F_k(b)$. Note that $v \in F_k(B) \subseteq G(B)$. We have

$$d\big(u, G(B)\big) \leq d(u, v) = d\big(F_k(a), F_k(b)\big) \leq c_k d(a, b) \leq c\, d(a, b) = c\, d(a, B) \leq c\, D(A, B).$$

It follows that $\max\limits_{u \in G(A)} d\big(u, G(B)\big) \leq c\, D(A, B)$. A similar argument (reversing the roles of A and B) shows that $\max\limits_{v \in G(B)} d\big(v, G(A)\big) \leq c\, D(A, B)$, and hence $D\big(G(A), G(B)\big) \leq c\, D(A, B)$, as required. This completes the proof of Part (1), and Part (2) follows immediately from the Banach Fixed Point Theorem (and its proof), because $\mathcal{K}$ is complete.

7.29 Definition: A (commutative) **algebra** over a field F is a vector space U with a binary multiplication operation such that for all $u,v,w\in U$ and all $t\in F$ we have $uv=vu$, $u(v+w)=uv+uw$, and $(tu)v=t(uv)$. A subspace $A\subseteq U$ is a **subalgebra** of U when it is an algebra using (the restriction of) the same operations used in U. Verify that a subset $A\subseteq U$ is a subalgebra of U when $0\in A$ and for all $u,v\in A$ and all $t\in F$ we have $tu\in A$, $u+v\in A$ and $uv\in A$.

7.30 Example: When X is a metric space, the vector space $\mathcal{F}(X)$ of all functions $f:X\to\mathbb{R}$ is an algebra over $\mathbb{R}$, and $\mathcal{B}(X)$, $\mathcal{C}(X)$, and $\mathcal{C}_b(X)$ are all subalgebras.

7.31 Example: When $a\leq b$, the space $\mathcal{P}[a,b]$ of polynomial maps $f:[a,b]\to\mathbb{R}$ and the space $\mathcal{C}^1[a,b]$ of continuously differentiable maps are subalgebras of the algebra $\mathcal{C}[a,b]$ of continuous maps $f:[a,b]\to\mathbb{R}$, and the space $\mathcal{R}[a,b]$ of Riemann integrable functions is a subalgebra of the algebra $\mathcal{B}[a,b]$ of bounded functions $f:[a,b]\to\mathbb{R}$.

7.32 Example: Show that $f(x)=|x|$ lies in the closure of $\mathcal{P}[-1,1]$ in $\big(\mathcal{C}[-1,1],d_\infty\big)$.

Solution: Let $\epsilon>0$ and let $a=\frac{\epsilon}{2}$. Let $g(x)=\sqrt{x+a^2}$ and let $p_n(x)$ be the n^{th} Taylor polynomial for $g(x)$ centred at 1: to be explicit, for $\left|\frac{x-1}{1+a^2}\right|$ we have

$$g(x)=\big((x-1)+(1+a^2)\big)^{1/2}=\sqrt{1+a^2}\,\left(1+\tfrac{x-1}{1+a^2}\right)^{1/2}=\sqrt{1+a^2}\,\sum_{k=1}^{\infty}\binom{1/2}{k}\left(\tfrac{x-1}{1+a^2}\right)^k,$$

and we have

$$p_n(x)=\sqrt{1+a^2}\,\sum_{k=0}^{n}\binom{1/2}{k}\left(\tfrac{x-1}{1+a^2}\right)^k.$$

Note that $p_n\to g$ pointwise for $\left|\frac{x-1}{1+a^2}\right|<1$, that is for all $x\in(-a^2,2+a^2)$, and $f_n\to g$ uniformly on $[0,2]$ (hence also on $[0,1]$). Choose $n\in\mathbb{Z}^+$ such that $\big|p_n(x)-g(x)\big|<a=\frac{\epsilon}{2}$ for all $x\in[0,1]$. Also note that for all $x\in\mathbb{R}$ we have

$$\Big||x|-g(x^2)\Big|=\sqrt{x^2+a^2}-\sqrt{x^2}=\frac{a^2}{\sqrt{x^2+a^2}+\sqrt{x^2}}\leq a=\tfrac{\epsilon}{2},$$

so for all $x\in[-1,1]$, we have $x^2\in[0,1]$, hence

$$\Big||x|-p_n(x^2)\Big|\leq\Big||x|-g(x^2)\Big|+\Big|g(x^2)-p_n(x^2)\Big|<\tfrac{\epsilon}{2}+\tfrac{\epsilon}{2}=\epsilon.$$

7.33 Definition: Let $A\subseteq\mathcal{C}(X)$. We say that A **separates points** when for all $x,y\in X$ with $x\neq y$ there exist $f\in A$ with $f(x)\neq f(y)$. We say that A **vanishes nowhere** when for all $a\in X$ there exists $f\in A$ such that $f(x)\neq 0$. Note that if $1\in A$ (where 1 denotes the constant function) the A vanishes nowhere.

7.34 Theorem: *(The Stone-Weierstrass Theorem) Let X be a compact metric space and let $A \subseteq \mathcal{C}(X)$ be an algebra. If A separates points and vanishes nowhere then $\overline{A} = \mathcal{C}(X)$ (using the supremum metric d_∞).*

Proof: Note first that $\overline{A}$ is also a subalgebra of $\mathcal{C}(X)$. Indeed given $f, g \in \overline{A}$ and $c \in \mathbb{R}$, we can choose sequences (f_n) and (g_n) in A such that $f_n \to f$ and $g_n \to g$ in $\mathcal{C}(X)$ (that is $f_n \to f$ and $g_n \to g$ uniformly on X), and then we have $cf_n \to cf$, $f_n + g_n \to f + g$ and $f_n g_n \to fg$ uniformly on X, and hence $cf \in \overline{A}$, $f + g \in \overline{A}$ and $fg \in \overline{A}$. Also note that $\overline{A}$ separates points and vanishes nowhere, and so we may assume, without loss of generality, that A is closed.

Next we claim that if $f \in A$ then we also have $|f| \in A$. Let $f \in A \subseteq \mathcal{C}(X)$. Choose $m > 0$ with $m \geq \|f\|_\infty$. Let $g = \frac{1}{m} f$ and note that $g \in A$ with $\|g\|_\infty \leq 1$, that is $g(x) \in [-1, 1]$ for all $x \in X$. Let $\epsilon > 0$. By Example 6.17, we can choose a polynomial $p_0(x) = a_0 + a_1 x + \cdots + a_n x^n$ such that $\big|p_0(u) - |u|\big| \leq \frac{\epsilon}{2}$ for all $u \in [-1, 1]$. Let $p(x) = p_0(x) - a_0$ and note that $\big|p(u) - |u|\big| \leq \epsilon$ for all $u \in [-1, 1]$. For all $x \in X$, we have $g(x) \in [-1, 1]$ and so $\big|p\big(g(x)\big) - |g(x)|\big| < \epsilon$. Note that the function $h(x) = p\big(g(x)\big) = a_1 g(x) + a_2 g(x)^2 + \cdots a_n g(x)^n$ lies in A (because $g \in A$ and A is an algebra). This shows that for every $\epsilon > 0$ we can find $h \in A$ with $\big|h - |g|\big| < \epsilon$, and (since A is closed) it follows that $|g| \in A$ and hence $|f| = m|g| \in A$.

Next we note that if $f, g \in A$ then we also have $\max\{f, g\} \in A$ and $\min\{f, g\} \in A$ because

$$\max\{f, g\} = \frac{f + g}{2} + \frac{|f - g|}{2} \quad \text{and} \quad \min\{f, g\} = \frac{f + g}{2} - \frac{|f - g|}{2}$$

and it follows, inductively, that if $f_1, f_2, \cdots, f_n \in A$ then we have $\max\{f_1, \cdots, f_n\} \in A$ and $\min\{f_1, \cdots, f_n\} \in A$.

We claim that for all $r, s \in \mathbb{R}$ and for all $a, b \in X$ with $a \neq b$, there is a function $g \in A$ with $g(a) = r$ and $g(b) = s$. Let $r, s \in \mathbb{R}$ and let $a, b \in X$ with $a \neq b$. Since A separates points, we can choose $h \in A$ with $h(a) \neq h(b)$. Since A vanishes nowhere, we can choose $k, \ell \in A$ with $k(a) \neq 0$ and $\ell(b) \neq 0$. Define $u, v \in A$ by

$$u(x) = \big(h(x) - h(b)\big)k(x) \quad \text{and} \quad v(x) = \big(h(a) - h(x)\big)\ell(x)$$

and note that $u(a) \neq 0$ and $u(b) = 0$ while $v(a) = 0$ and $v(b) \neq 0$. Then define $g \in A$ by

$$g(x) = r\,\frac{u(x)}{u(a)} + s\,\frac{v(x)}{v(b)}$$

to obtain $g(a) = r$ and $g(b) = s$, as required.

We claim that for every $f \in \mathcal{C}(X)$, for every $a \in X$ and for every $\epsilon > 0$, there is a function $h \in A$ such that $h(a) = f(a)$ and $h(x) < f(x) + \epsilon$ for all $x \in X$. Let $f \in \mathcal{C}(X)$, let $a \in X$ and let $\epsilon > 0$. For each $b \in X$, by the previous claim we can choose $g_b \in A$ such that $g_b(a) = f(a)$ and $g_b(b) = f(b)$. For each $b \in X$, since f and g_b are continuous at b, we can choose $r_b > 0$ such that for all $x \in B(b, r_b)$ we have

$$\big|f(x) - f(b)\big| < \tfrac{\epsilon}{2} \quad \text{and} \quad \big|g_b(x) - g_b(b)\big| < \tfrac{\epsilon}{2}\ \text{, hence } \big|g_b(x) - f(x)\big| < \epsilon.$$

Since X is compact and the set $\big\{B(b, r_b)\,\big|\,b \in X\big\}$ covers X, we can choose $b_1, b_2, \cdots, b_n \in X$ such that $X = \bigcup\limits_{k=1}^{n} B(b_k, r_{b_k})$, and then we let

$$h = \min\big\{g_{b_1}, g_{b_2}, \cdots, g_{b_n}\big\} \in A.$$

For all $x \in X$ we can choose an index k such that $x \in B(b_k, r_{a_k})$ and then we have $h(x) \leq g_{b_k}(x) < f(x) + \epsilon$, as required.

Finally, we complete the proof by showing that for every $f \in \mathcal{C}[0,1]$ and every $\epsilon > 0$ there exists $g \in A$ such that $|g(x) - f(x)| < \epsilon$ for all $x \in X$. Let $f \in \mathcal{C}(X)$ and let $\epsilon > 0$. For each $a \in X$, by the previous claim we can choose $h_a \in A$ such that $h_a(a) = f(a)$ and $h_a(x) < f(x) + \epsilon$ for all $x \in X$. For each $a \in X$, since f and h_a iare continuous at a, we can choose $s_a > 0$ such that for all $x \in B(a, s_a)$ we have

$$\left|f(x) - f(a)\right| < \tfrac{\epsilon}{2} \;\text{ and }\; \left|h_a(x) - h_a(a)\right| < \tfrac{\epsilon}{2} \;\text{ hence }\; \left|h_a(x) - f(x)\right| < \epsilon.$$

Since X is compact and $\left\{B(a_k, s_k) \,\middle|\, a \in X\right\}$ covers X, we can choose $a_1, a_2, \cdots, a_m \in X$ such that $X = \bigcup\limits_{k=1}^{m} B(a_k, s_{a_k})$, and then we chose

$$g = \max\left\{h_{a_1}, h_{a_2}, \cdots, h_{a_m}\right\} \in A.$$

For all $x \in X$ we can choose an index k such that $x \in B(a_k, s_{a_k})$ and we can choose an index ℓ such that $g(x) = h_{a_\ell}(x)$ and then we have

$$g(x) \geq h_{a_k}(x) > f(x) - \epsilon \;\text{ and }\; g(x) = h_{a_\ell}(x) < f(x) + \epsilon.$$

7.35 Corollary: *(The Weierstrass Approximation Theorem) Let $X \subseteq \mathbb{R}^n$ be compact and let $f \in \mathcal{C}(X)$. Then for all $\epsilon > 0$ there exists a polynomial p in n variables such that $\left|p(x) - f(x)\right| < \epsilon$ for all $x \in X$.*

Proof: Each polynomial p in n-variables determines a continuous function $p : X \to \mathbb{R}$. The set $\mathcal{P}(X)$ of such polynomial functions is a subalgebra of $\mathcal{C}(X)$ which separates points and vanishes nowhere, so $\mathcal{P}(X)$ is dense in $\mathcal{C}(X)$, using the metric d_∞. This means that given $f \in \mathcal{C}(X)$, for all $\epsilon > 0$ we can choose $p \in \mathcal{P}(X)$ such that $\|p - f\|_\infty < \epsilon$, and hence $\left|p(x) - f(x)\right| < \epsilon$ for all $x \in X$.

7.36 Exercise: Let $A = \left\{ \sum\limits_{k=1}^{n} f_k(x) g_k(y) \,\middle|\, n \in \mathbb{Z}^+, f_k, g_k \in \mathcal{C}[0,1] \right\}$. Show that A is dense in $\mathcal{C}\big([0,1] \times [0,1]\big)$, using the metric d_∞.

7.37 Exercise: Let $A = \left\{ b_0 + \sum\limits_{k=1}^{n} (a_k \sin(kx) + b_k \cos(kx)) \,\middle|\, n \in \mathbb{Z}^+, a_k, b_k \in \mathbb{R} \right\}$. Show that for all $r \in [0, 2\pi]$, A is dense in $\mathcal{C}[0,r]$ but A is not dense in $\mathcal{C}[0,2\pi]$, using d_∞.

7.38 Definition: Let X be a set and let $S \subseteq \mathcal{F}(X) = \mathcal{F}(X, \mathbb{R})$. We say that S is **pointwise bounded** when for every $x \in X$ there exists $m = m(x) > 0$ such that $|f(x)| \leq m$ for every function $f \in S$. We say that S is **uniformly bounded** when there exists $m > 0$ such that $|f(x)| \leq m$ for every $x \in X$ and every $f \in S$.

Let X be a metric space and let $S \subseteq \mathcal{C}(X) = \mathcal{C}(X, \mathbb{R})$. We say that S is **equicontinuous** when for every $\epsilon > 0$ there exists $\delta > 0$ such that for every $f \in S$ and for all $x, y \in X$, if $d(x, y) < \delta$ then $d\big(f(x), f(y)\big) < \epsilon$

7.39 Note: When X is a compact metric space, by the Extreme Value Theorem, every continuous function $f : X \to \mathbb{R}$ is also bounded, so we have $\mathcal{C}(X) = \mathcal{C}_b(X)$, which is a complete metric space using the supremum norm. Unless otherwise stated, when we refer to the metric space $\mathcal{C}(X)$ it is understood that we are using the supremum metric.

7.40 Note: When X is a compact metric space and $S \subseteq \mathcal{C}(X)$, note that S is uniformly bounded if and only if S is bounded as a subspace of the metric space $\mathcal{C}(X)$.

7.41 Theorem: *Let X be a compact metric space and let (f_n) be a sequence in $\mathcal{C}(X)$. If the sequence (f_n) converges in the metric space $\mathcal{C}(X)$ (equivalently, if the sequence (f_n) converges uniformly on X) then the set $\{f_n\}$ is equicontinuous.*

Proof: Suppose (f_n) converges in $\mathcal{C}(X)$. Let $\epsilon > 0$. Since (f_n) converges in $\mathcal{C}(X)$ we can choose $\ell \in \mathbb{Z}^+$ such that for all $n, m \geq \ell$ we have $\|f_n - f_m\|_\infty < \frac{\epsilon}{3}$. Since X is compact, each of the functions f_n is uniformly continuous on X. Choose $\delta > 0$ such that for all $x, y \in X$ with $d(x, y) < \delta$ we have $\big|f_n(x) - f_n(y)\big| < \epsilon$ for each $n < \ell$ and we have $\big|f_\ell(x) - f_\ell(y)\big| < \frac{\epsilon}{3}$. Then for all $n \geq \ell$ and all $x, y \in X$ with $d(x, y) < \delta$ we have

$$\big|f_n(x) - f_n(y)\big| \leq \big|f_n(x) - f_\ell(x)\big| + \big|f_\ell(x) - f_\ell(y)\big| + \big|f_\ell(y) - f_n(y)\big| < \epsilon.$$

7.42 Corollary: *Let X be a compact metric space. Then every compact set $S \subseteq \mathcal{C}(X)$ is equicontinuous.*

Proof: Let $S \subseteq \mathcal{C}(X)$. Suppose that S is not equicontinuous. Choose $\epsilon > 0$ such that for all $\delta > 0$ there exists $f \in S$ and there exist $x, y \in X$ with $d(x, y) < \delta$ such that $\big|f(x) - f(y)\big| \geq \epsilon$. For each $n \in \mathbb{Z}^+$, choose $f_n \in S$ such that there exist $x, y \in X$ with $d(x, y) < \frac{1}{2^n}$ such that $\big|f_n(x) - f_n(y)\big| \geq \epsilon$. Then no subsequence of (f_n) can possibly converge in S (using the supremum metric) and so S cannot be compact.

7.43 Theorem: *Let X be a compact metric space and let (f_n) be a sequence in $\mathcal{C}(X)$. If the set $\{f_n\}$ is pointwise bounded and equicontinuous then the set $\{f_n\}$ is uniformly bounded and the sequence (f_n) has a convergent subsequence in $\mathcal{C}(X)$.*

Proof: Suppose that the set $\{f_n\}$ is pointwise bounded and equicontinuous. We claim that the set $\{f_n\}$ is uniformly bounded. Since $\{f_n\}$ is equicontinuous, we can choose $\delta > 0$ such that for all $n \in \mathbb{Z}^+$ and for all $x, y \in X$ with $d(x, y) < \delta$ we have $\big|f_n(x) - f_n(y)\big| < 1$. Since X is compact, we can choose $a_1, a_2, \cdots, a_\ell \in X$ such that $X = B(a_1, \delta) \cup \cdots \cup B(a_\ell, \delta)$. Since $\{f_n\}$ is pointwise bounded, we can choose $m > 0$ such that for each index k with $1 \leq k \leq \ell$ we have $\big|f_n(a_k)\big| \leq m$. Let $n \in \mathbb{Z}^+$ and $x \in X$. Choose an index k with $1 \leq k \leq \ell$ such that $x \in B(a_k, \delta)$. Since $d(x, a_k) < \delta$ we have $\big|f_n(x) - f_n(a_k)\big| < 1$ and so $\big|f_n(x)\big| \leq \big|f_n(x) - f_n(a_k)\big| + \big|f_n(a_k)\big| < 1 + m$. Thus the set $\{f_n\}$ is uniformly bounded, as claimed.

It remains to show that the sequence (f_n) has a convergent subsequence in $\mathcal{C}(X)$. Since X is compact, and hence separable, we can choose a countable dense subset $A \subseteq X$, say $A = \{a_1, a_2, a_3, \cdots\}$. We claim that the sequence $(f_n)_{n\geq 1}$ has a subsequence $(f_{n_k})_{k\geq 1}$ which converges pointwise on A. Since the real-valued sequence $\big(f_n(a_1)\big)_{n\geq 1}$ is bounded, we can choose a subsequence, which we shall write as $\big(f_{1,k}\big)_{k\geq 1} = \big(f_{1,1}, f_{1,2}, f_{1,3}, \cdots\big)$, of the sequence of functions $(f_n)_{n\geq 1}$ such that the real-valued sequence $\big(f_{1,k}(a_1)\big)_{k\geq 1}$ converges. Since the real-valued sequence $\big(f_{1,k}(a_2)\big)_{k\geq 1}$ is bounded, we can choose a subsequence $\big(f_{2,k}\big)$ of the sequence of functions $\big(f_{1,k}\big)$ such that the real-valued sequence $\big(f_{2,k}(a_2)\big)$ converges. Note that since $\big(f_{2,k}(a_1)\big)$ is a subsequence of the convergent sequence $\big(f_{1,k}(a_1)\big)$, it also converges. By recursively repeating this procedure, we construct sequences $(f_{n,k})_{k\geq 1}$ for each $n \geq 1$, such that $\big(f_{n+1,k}\big)_{k\geq 1}$ is a subsequence of $\big(f_{n,k}\big)_{k\geq 1}$ and the real-valued sequences $\big(f_{n,k}(a_j)\big)_{k\geq 1}$ converge for all j with $1 \leq j \leq n$. Let $\big(f_{n_k}\big)_{k\geq 1}$ denote the sequence $\big(f_{1,1}, f_{2,2,}, f_{3,3,}, \cdots\big)$, note that this is a subsequence of the original sequence (f_n), and the real-valued sequences $\big(f_{n_k}(a_j)\big)_{k\geq 1}$ converge for all indices $j \in \mathbb{Z}^+$, so the subsequence (f_{n_k}) converges pointwise on A, as required.

Finally, we claim that the above subsequence (f_{n_k}) converges in $\mathcal{C}(X)$. Let $\epsilon > 0$. Since the set $\{f_n\}$ is equicontinuous we can choose $\delta > 0$ such that for all $n \in \mathbb{Z}^+$ and all $x, y \in X$ with $d(x, y) < \delta$ we have $\big|f_n(x) - f_n(y)\big| < \frac{\epsilon}{3}$. Since A is dense in X, the set $\mathcal{U} = \big\{B(a_n, \delta)\big|n \in]Z^+\big\}$ is an open cover of X. Since X is compact, we can choose a finite subcover of $\mathcal{U}$, so we can choose $a_1, a_2, \cdots, a_p \in X$ such that $X = B(a_1, \delta) \cup \cdots \cup B(a_p, \delta)$. Since the sequences $\big(f_{n_k}(a_j)\big)_{k\geq 1}$ all converge, we can choose $m \in \mathbb{Z}^+$ such that for all $j \in \mathbb{Z}^+$ with $1 \leq j \leq p$ and all $k, \ell \in \mathbb{Z}^+$ with $k, \ell \geq m$ we have $\big|f_{n_k}(a_j) - f_{n_\ell}(a_j)\big| < \frac{\epsilon}{3}$. Let $x \in X$ and let $k, \ell \in \mathbb{Z}^+$ with $k, \ell \geq m$. Choose an index j with $1 \leq j \leq p$ such that $x \in B(a_j, \delta)$. Then we have

$$\big|f_{n_k}(x) - f_{n_\ell}(x)\big| \leq \big|f_{n_k}(x) - f_{n_k}(a_j)\big| + \big|f_{n_k}(a_j) - f_{n_\ell}(a_j)\big| + \big|f_{n_\ell}(a_j) - f_{n_\ell}(x)\big| < \epsilon.$$

7.44 Theorem: *(The Arzela-Ascoli Theorem) Let X be a compact metric space and let $S \subseteq \mathcal{C}(X)$, using the supremum metric.*

(1) S is compact if and only if S is closed, pointwise bounded, and equicontinuous.
(2) If S is pointwise bounded and equicontinuous, then $\overline{S}$ is compact.

Proof: To prove Part 1, suppose that S is compact. Then we know that S is closed and bounded and we know (from Corollary 6.9) that S is equicontinuous. Since S is bounded, using the supremum metric, it follows that S is uniformly bounded, hence also pointwise bounded.

Suppose, conversely, that S is closed, pointwise bounded, and equicontinuous. Let (f_n) be a sequence in S. Since S is pointwise bounded and equicontinuous, the subset $\{f_n\}$ is also pointwise bounded and equicontinuous. By the above theorem, the sequence (f_n) has a convergent subsequence (f_{n_k}) in $\mathcal{C}(X)$. Since S is closed, the limit of this subsequence lies in S. This proves that every sequence in S has a subsequence which converges in S, and so S is compact.

This completes the proof of Part 1, and we leave the proof of Part 2 as an exercise.

7.45 Theorem: *(Peano) Let $U \in \mathbb{R}^2$ be open, let $(a,b) \in U$, and let $F : U \to \mathbb{R}$ be continuous. Then there exists $d > 0$ such that the differential equation $\frac{dy}{dx} = F(x,y)$ has a solution $y = f(x)$ which is defined for all $x \in [a-d, a+d]$.*

Proof: Choose a closed rectangle Q with $(a,b) \in Q \subseteq U$. Since Q is compact, $\big|F(x,y)\big|$ attains its maximum value on Q, let $M = \max\big\{|F(x,y)| \,\big|\, (x,y) \in Q\big\}$. Choose $d > 0$ so that $R = [a-d, a+d] \times [b-Md, b+Md] \subseteq Q$.

Fix $n \in \mathbf{Z}^+$. Since R is compact so that F is uniformly continuous on R, we can choose $\delta > 0$ so that for all $(x_1,y_1),(x_2,y_2) \in R$,

$$\big|(x_1,y_1) - (x_2,y_2)\big| < \delta \Longrightarrow \big|F(x_1,y_1) - F(x_2,y_2)\big| \le \tfrac{1}{n}.$$

Choose $\ell \in \mathbb{Z}^+$ so $\frac{d}{\ell} < \frac{\delta}{M+1}$ and let $c_k = a + \frac{kd}{\ell}$ for $0 \le k < \ell$ so $a = c_0 < c_1 < \cdots < c_\ell = a+d$ with $c_{k+1} - c_k = \frac{d}{\ell} < \frac{\delta}{M+1}$ for all $0 \le k < \ell$. Let $f = f_n : [a, a+d] \to \mathbb{R}$ be the continuous, piecewise linear function with $f(a) = b$ such that $f'(x) = F\big(c_k, f(c_k)\big)$ for all $t \in (c_k, c_{k+1})$ (the function $f = f_n$ is constructed recursively by beginning with $f(a) = b$ and then, having defined $f(x)$ for all $x \in [a, c_k]$, define $f(x) = f(c_k) + F\big(c_k, f(c_k)\big)(x - c_k)$ for all $x \in [c_k, c_{k+1}]$).

Claim 1: we claim that for all $x_1, x_2 \in [a, a+d]$ we have

$$\big|f(x_1) - f(x_2)\big| \le M|x_1 - x_2|.$$

Let $x_1, x_2 \in [a, a+d]$ with $x_1 \le x_2$. For $0 \le k < p$, let $m_k = F(c_k, f(c_k))$ and note that $|m_k| = \big|F(c_k, f(c_k))\big| \le M$ for all k. When $x_1, x_2 \in [c_k, c_{k+1}]$ with $x_1 \le x_2$, we have $f(x_2) = f(x_1) + m_k(x - c_k)$ so that $\big|f(x_2) - f(x_1)\big| = \big|m_k(x_2 - x_1)\big| \le M(x_2 - x_1)$, and when $x_1 \in [c_j, c_{j+1}]$ and $x_2 \in [c_k, c_{k+1}]$ with $j < k$ we have

$$f(x_2) = f(x_1) + m_j(c_{j+1} - x_1) + m_{j+1}(c_{j+2} - c_{j+1}) + \cdots + m_{k-1}(c_k - c_{k-1}) + m_k(x_2 - c_k)$$

so $\big|f(x_2) - f(x_1)\big| \le M(c_{j+1} - x_1) + M(c_{j+2} - c_{j+1}) + \cdots + M(c_k - c_{k-1}) + M(x_2 - c_k)$, that is $\big|f(x_2) - f(x_1)\big| \le M(x_2 - x_1)$, as required.

Claim 2: we claim that when $x \in [c_k, c_{k+1}]$ we have

$$\int_{c_k}^{x} \Big|m_k - F\big(t, f(t)\big)\Big|\, dt \le \tfrac{d}{n}.$$

Let $x \in [c_k, c_{k+1}]$ and let $t \in [c_k, x]$. Then $|c_k - t| \le c_{k+1} - c_k = \frac{d}{\ell} < \frac{\delta}{M+1}$. By Claim 1, we have $\big|f(c_k) - f(t)\big| < \frac{M\delta}{M+1}$, so

$$\big|(c_k, f(c_k)) - (t, f(t))\big| \le |c_k - t| + \big|f(c_k) - f(t)\big| < \tfrac{\delta}{M+1} + \tfrac{M\delta}{M+1} = \delta.$$

By the choice of δ, we have $\big|m_k - F\big(t, f(t)\big)\big| = \big|F\big(c_k, f(c_k)\big) - F\big(t, f(t)\big)\big| \le \frac{1}{n}$. This holds for all $t \in [c_k, x]$, so

$$\int_{c_k}^{x} \Big|m_k - F\big(t, f(t)\big)\Big|\, dt \le \int_{c_k}^{x} \tfrac{1}{n}\, dt = \tfrac{1}{n}\,(x - c_k) \le \tfrac{d}{n}$$

as claimed.

Claim 3. we claim that for all $x \in [a, a+d]$ we have

$$\left|\big(f(x) - f(a)\big) - \int_a^x F\big(t, f(t)\big)\, dt\right| \le \tfrac{d}{n}.$$

Let $x \in [a, a+d]$. Choose an index k so that $x \in [c_k, c_{k+1}]$. Then

$$f(x) - f(a) = \sum_{j=0}^{k-1} m_j(c_{j+1} - c_j) + m_k(x - c_k) = \sum_{j=0}^{k-1} \int_{c_j}^{c_{j+1}} m_j\, dt + \int_{c_k}^{x} m_k\, dt$$

and so,
$$\left|\big(f(x)-f(a)\big)-\int_a^x F\big(t,f(t)\big)\,dt\right|$$
$$=\left|\sum_{j=0}^{k-1}\int_{c_j}^{c_{j+1}}\Big(m_j-F\big(t,f(t)\big)\Big)\,dt+\int_{c_k}^{x}\Big(m_k-F\big(t,f(t)\big)\Big)\,dt\right|$$
$$\leq\sum_{j=0}^{k-1}\int_{c_j}^{c_{j+1}}\Big|m_j-F\big(t,f(t)\big)\Big|\,dt+\int_{c_k}^{x}\Big|m_k-F\big(F\big(t,f(t)\big)\Big|\,dt$$
$$\leq\sum_{j=0}^{k-1}\tfrac{1}{n}\,(c_{j+1}-c_j)+\tfrac{1}{n}\,(x-c_k)=\tfrac{1}{n}\,(x-c_0)=\tfrac{1}{n}\,(x-a)\leq\tfrac{d}{n}\,,$$
(where we used Claim 2 on the final line above), as claimed.

We repeat the above construction for every $n\in\mathbb{Z}^+$ to obtain a sequence of functions $(f_n)_{n\geq1}$. Each function f_n satisfies Claims 1, 2 and 3. Let $S=\big\{f_n\,\big|\,n\in\mathbb{Z}^+\big\}\subseteq\mathcal{C}[a,a{+}d]$. Note that by Claim 1, the set S is equicontinuous $\big($indeed given $\epsilon>0$, if $|x_1-x_2|<\frac{\epsilon}{M}$ then $\big|f_n(x_1)-f_n(x_2)\big|\leq M|x_1-x_2|<\epsilon\big)$ and the set S uniformly bounded $\big($indeed since $f_n(a)=b$ and $\big|f_n(x)-f_n(a)\big|\leq M|x-a|\leq Md$ we have $b-Md\leq f_n(x)\leq b+Md$ for all $x\big)$. By the Arzela-Ascoli Theorem, the closure $\overline{S}$ of S in $\big(\mathcal{C}[a,a{+}d],d_\infty\big)$ is compact. Thus we can choose a subsequence $(f_{n_k})_{k\geq1}$ of $(f_n)_{n\geq1}$ which converges in $\overline{S}\subseteq\mathcal{C}[a,a{+}d]$ using the metric d_∞, that is $(f_{n_k})_{k\geq1}$ converges uniformly on $[a,a{+}d]$ to some continuous function $g:[a,a{+}d]\to\mathbb{R}$.

Claim 4: we claim that the above map $g:[a,a{+}d]\to\mathbb{R}$ is a solution to the given differential equation. First we note that when $\big\|f-g\big\|_\infty<\delta$, for all $t\in[a,a{+}d]$ we have $\big|(t,f(t))-(t,g(t))\big|=\big|f(t)-g(t)\big|\leq\|f-g\|_\infty<\delta$ so that (by the choice of δ) we have $\big|F(t,f(t))-F(t,g(t))\big|\leq\frac{1}{n}$ and hence
$$\int_a^x\Big|F\big(t,f(t)\big)-F\big(t,g(t)\big)\Big|\,dt\leq\int_a^x\tfrac{1}{n}\,dt=\tfrac{1}{n}\,(x-a)\leq\tfrac{d}{n}.$$
Given $\epsilon>0$ we can choose $k\in\mathbb{Z}^+$ such that $\|f_{n_k}-g\|_\infty<\delta$ and $\|f_{n_k}-g\|_\infty<\frac{\epsilon}{3}$ and $\frac{1}{n_k}<\frac{\epsilon}{3d}$. Write $n=n_k$ and $f=f_n=f_{n_k}$. Then for all $x\in[a,a{+}d]$ we have
$$\left|\big(g(x)-g(a)\big)-\int_a^x F\big(t,g(t)\big)\,dt\right|\leq\Big|\big(g(x)-g(a)\big)-\big(f(x)-f(a)\big)\Big|$$
$$+\left|\big(f(x)-f(a)\big)-\int_a^x F\big(t,f(t)\big)\,dt\right|+\left|\int_a^x F\big(t,f(t)\big)\,dt-\int_a^x F\big(t,g(t)\big)\,dt\right|$$
$$\leq\Big|g(x)-f(x)\Big|+\left|\big(f(x)-f(a)\big)-\int_a^x F\big(t,f(t)\big)\,dt\right|+\int_a^x\Big|F\big(t,f(t)\big)-F\big(t,g(t)\big)\Big|\,dt$$
$$\leq\|f-g\|_\infty+\tfrac{d}{n}+\tfrac{d}{n}<\tfrac{\epsilon}{3}+\tfrac{\epsilon}{3}+\tfrac{\epsilon}{3}=\epsilon.$$
Since $\epsilon>0$ was arbitrary, it follows that for all $x\in[a,a{+}d]$
$$g(x)=g(a)+\int_a^x F\big(t,g(t)\big)\,dt\,.$$
By the Fundamental Theorem of Calculus, g is differentiable wih $g'(x)=F\big(x,g(x)\big)$ for all $x\in[a,a{+}d]$, and so g is a solution of the given differential equation, as claimed.

Finally, we repeat the above procedure to obtain a solution $g:[a{-}\,d,a]\to\mathbb{R}$ then join the two solutions to obtain a solution $g:[a{-}d,a{+}d]\to\mathbb{R}$.