2.3 Compactness

Recall that Bolzano-Weierstrass Theorem asserts that every sequence in a closed bounded interval has a convergent subsequence in this interval. The result still holds for all closed, bounded sets in $\mathbb{R}^n$. In general, a set $E \subset (X,d)$ is **compact** if every sequence has a convergent subsequence with limit in E. This property is also called **sequentially compact** to stress that the behavior of sequences is involved in the definition. The space (X,d) is called a **compact space** is X is a compact set itself. According to this definition, every interval of the form $[a,b]$ is compact in $\mathbb{R}$ and sets like $[a_1,b_1]\times[a_2,b_2]\times\cdots[a_n,b_n]$ and $\overline{B_r(x)}$ are compact in $\mathbb{R}^n$ under the Euclidean metric. In a general metric space, the notion of a bounded set makes perfect sense. Indeed, a set A is called a **bounded set** if there exists some ball $B_r(x)$ for some $x \in X$ and $r>0$ such that $A \subset B_r(x)$. Now we investigate the relation between a compact set and a closed bounded set. First of all, we have

Proposition 2.1. *Every compact set in a metric space is closed and bounded.*

Proof. Let K be a compact set. To show that it is closed, let $\{x_n\} \subset K$ and $x_n \to x_0$. We need to show that $x_0 \in K$. As K is compact, there exists a subsequence $\{x_{n_j}\} \subset K$ converging to some z in K. By the uniqueness of limit, we have $x_0 = z \in K$, so $x_0 \in K$ and K is closed.

Next, we show that K is bounded. If on the contrary it is not, for a fixed point w, K is not contained in the balls $B_n(w)$ for all n. Picking $x_n \in K \setminus B_n(w)$, we obtain a sequence $\{x_n\}$ satisfying $d(x_n,w)\to\infty$ as $n\to\infty$. By the compactness of K, there is a subsequence $\{x_{n_j}\}$ converging to some z in K. Therefore, for all sufficiently large n_j, $x_{n_j} \in B_1(z)$. By the triangle inequality,

$$\begin{aligned} d(x_{n_j},w) &\le d(x_{n_j},w)+d(w,z) \\ &\le 1+d(x_1,z)<\infty\ , \end{aligned}$$

contradicting $d(x_{n_j},w)\to\infty$ as $n_j\to\infty$. Hence K must be bounded. □

As a consequence of Bolzano-Weierstrass Theorem every sequence in a bounded and closed set in $\mathbb{R}^n$ contains a convergent subsequence. Thus a set in $\mathbb{R}^n$ is compact if and only if it is closed and bounded. Proposition 2.10 tells that every compact set is in general closed and bounded, but the converse is not always true. To describe an example we need to go beyond $\mathbb{R}^n$ where we can be free of the binding of Bolzano-Weierstrass Theorem. Consider the set $\mathcal{S} = \{f \in C[0,1] : 0 \le f(x) \le 1\}$. Clearly it is closed and bounded in $C[0,1]$. We claim that it is not compact. For, consider the sequence $\{f_n\}$ in $(C[0,1], d_\infty)$ given by

$$f_n(x) = \begin{cases} nx, & x \in [0,\frac{1}{n}] \\ 1, & x \in [\frac{1}{n},1]. \end{cases}$$

$\{f_n(x)\}$ converges pointwisely to the function $f(x) = 1$, $x \in (0,1]$ and $f(0) = 0$ which is discontinuous at $x = 0$, that is, f does not belong to $C[0,1]$. If $\{f_n\}$ has a convergent subsequences, then it must converge uniformly to f. But this is impossible because the uniform limit of a sequence of continuous functions must be continuous. Hence $\mathcal{S}$ cannot be compact. In fact, a remarkable theorem in functional analysis asserts that the closed unit ball in a normed space is compact if and only if the normed space if and only if the normed space is of finite dimension.

Since convergence of sequences can be completely described in terms of open/closed sets, it is natural to attempt to describe the compactness of a set in terms of these new notions. The answer to this challenging question is a little strange at first sight. Let us recall the following classical result:

Heine-Borel Theorem. *Let $\{I_j\}_{j=1}^{\infty}$ be a family of open intervals satisfying*

$$[a,b] \subset \bigcup_{j=1}^{\infty} I_j \ .$$

There is always a finite subfamily $\{I_{j_1}, \cdots, I_{j_K}\}$ such that

$$[a,b] \subset \bigcup_{k=1}^{K} I_{j_k} \ .$$

This property is not true for open (a,b). Indeed, the intervals $\{(a+1/j, b-1/j)\}$ satisfy $(a,b) \subset \bigcup_j (a+1/j, b-1/j)$ but there is no finite subcover. It is a good exercise to show that Heine-Borel Theorem is equivalent to Bolzano-Weierstrass Theorem. (When I was an undergraduate in this department in 1974, we were asked to show this equivalence, together with the so-called Nested Interval Theorem, in a MATH2050-like course.) This equivalence motivates how to describe compactness in terms of the language of open/closed sets.

We introduce some terminologies. First of all, an **open cover** of a subset E in a metric space (X,d) is a collection of open sets $\{G_\alpha\}, \alpha \in \mathcal{A}$, satisfying $E \subset \bigcup_{\alpha \in \mathcal{A}} G_\alpha$. A set $E \subset X$ satisfies the **finite cover property** if whenever $\{G_\alpha\}, \alpha \in \mathcal{A}$, is an open cover of E, there exist a subcollection consisting of finitely many $G_{\alpha_1}, \ldots, G_{\alpha_N}$ such that $E \subset \bigcup_{j=1}^N G_{\alpha_j}$. ("Every open cover has a finite subcover.") A set E satisfies the **finite intersection property** if whenever $\{F_\alpha\}, \alpha \in \mathcal{A}$, are relatively closed sets in E satisfying $\bigcap_{j=1}^N F_{\alpha_j} \neq \phi$ for any finite subcollection F_{α_j}, $\bigcap_{\alpha \in \mathcal{A}} F_\alpha \neq \phi$. Here a set $F \subset E$ is relatively closed means F is closed in the subspace E. We know that it implies $F = A \cap E$ for some closed set A. Therefore, when E is closed, a relatively closed subset is also closed. (In this semester I have skipped the discussion on subspace and relative openness and closedness. You may simply ignore the difference between a closed set and a relatively closed set in these notes.)

Proposition 2.2. *A closed set has the finite cover property if and only if it has the finite intersection property.*

Proof. Let E be a non-empty closed set in (X, d).

$\Rightarrow$) Suppose $\{F_\alpha\}$, F_α closed sets contained in E, satisfies $\bigcap_{j=1}^N F_{\alpha_j} \neq \phi$ for any finite subcollection but $\bigcap_{\alpha \in \mathcal{A}} F_\alpha = \phi$. As E is closed, each F_α is closed in X, and

$$E = E \setminus \bigcap_{\alpha \in \mathcal{A}} F_\alpha = \bigcup_{\alpha \in \mathcal{A}} (E \cap F'_\alpha) \subset \bigcup_{\alpha \in \mathcal{A}} F'_\alpha.$$

By the finite covering property we can find $\alpha_1, \ldots, \alpha_N$ such that $E \subset \bigcup_{j=1}^N F'_{\alpha_j}$, but then $\phi = E \setminus E \supset E \setminus \bigcup_1^N F'_{\alpha_j} = \bigcap_{j=1}^N F_{\alpha_j}$, contradiction holds.

$\Leftarrow$) If $E \subset \bigcup_{\alpha \in \mathcal{A}} G_\alpha$ but $E \subsetneq \bigcup_{j=1}^N G_{\alpha_j}$ for any finite subcollection of $\mathcal{A}$, then

$$\phi \neq E \setminus \bigcup_{j=1}^N G_{\alpha_j} = \bigcap_{j=1}^N \left(E \setminus G_{\alpha_j}\right)$$

which implies $\bigcap_{\alpha \in \mathcal{A}} \left(E \setminus G_\alpha\right) \neq \phi$ by the finite intersection property. Note that each $E \setminus G_{\alpha_j}$ is closed. Using $E \bigcap \left(\bigcup_{\alpha \in \mathcal{A}} G_\alpha\right)' = \bigcap_{\alpha \in \mathcal{A}} \left(E \setminus G_\alpha\right)$, we have $E \subsetneq \bigcup_{\alpha \in \mathcal{A}} G_\alpha$, contradicting our assumption.

□

Remark 2.1. It is no hard to see that every closed subset of a closed set satisfies the finite cover property/finite intersection property if the closed set itself satisfies the same property.

Proposition 2.3. *Let E be compact in a metric space. For each $\alpha > 0$, there exist finitely many balls $B_\alpha(x_1), \ldots, B_\alpha(x_N)$ such that $E \subset \bigcup_{j=1}^N B_\alpha(x_j)$ where $x_j, 1 \leq j \leq N$, are in E.*

Proof. Pick $B_\alpha(x_1)$ for some $x_1 \in E$. Suppose $E \setminus B_\alpha(x_1) \neq \phi$. We can find $x_2 \notin B_\alpha(x_1)$ so that $d(x_2, x_1) \geq \alpha$. Suppose $E \setminus \left(B_\alpha(x_1) \bigcup B_\alpha(x_2)\right)$ is non-empty. We can find $x_3 \notin B_\alpha(x_1) \bigcup B_\alpha(x_2)$ so that $d(x_j, x_3) \geq \alpha$, $j = 1, 2$. Keeping this procedure, we obtain a sequence $\{x_n\}$ in E such that

$$E \setminus \bigcup_{j=1}^n B_\alpha(x_j) \neq \phi \quad \text{and} \quad d(x_j, x_n) \geq \alpha,\ j = 1, 2, \ldots, n-1.$$

By the compactness of E, there exists $\left\{x_{n_j}\right\}$ and $x \in E$ such that $x_{n_j} \to x$ as $j \to \infty$. But then $d(x_{n_j}, x_{n_k}) < d(x_{n_j}, x) + d(x_{n_k}, x) \to 0$, contradicting $d(x_j, x_n) \geq \alpha$ for all $j < n$. Hence one must have $E \setminus \bigcup_{j=1}^N B_\alpha(x_j) = \phi$ for some finite N.

□

Sometimes the following terminology is convenient. A set E is called **totally bounded** if for each $\varepsilon > 0$, there exist $x_1, \cdots, x_n \in X$ such that $E \subset \cup_{k=1}^n B_\varepsilon(x_k)$. Proposition 2.12 simply states that every compact set is totally bounded. We will use this property of a compact set again in the next chapter.

Theorem 2.4. *Let E be a closed set in (X, d). The followings are equivalent:*

(a) E is compact;

(b) E satisfies the finite cover property; and

(c) E satisfies the finite intersection property.

Proof. The equivalence between (b) and (c) has been established in Proposition 2.11.

(a)$\Rightarrow$ (b). Let $\{G_\alpha\}$ be an open cover of E without finite subcover and we will draw a contradiction. By Proposition 2.12, for each $k \geq 1$, there are finitely many balls of radius $1/k$ covering E. We can find a set $B_{1/k} \cap E$ (suppress the irrelevant center) which cannot be covered by finitely many members in $\{G_\alpha\}$. Pick $x_k \in B_{1/k} \cap E$ to form a sequence. By the compactness of E, we can extract a subsequence $\{x_{k_j}\}$ such that $x_{k_j} \to x$ for some $x_0 \in E$. Since $\{G_\alpha\}$ covers E, there must be some G_{α_1} that contains x_0. As G_{α_1} is open and the radius of B_{1/k_j} tends to 0, we deduce that, for all sufficiently large k_j, $B_{1/k_j} \cap E$ is contained in G_{α_1}. In other words, G_{α_1} forms a single subcover of $B_{1/k} \cap E$, contradicting our choice of $B_{1/k_j} \cap E$. Hence (b) must be valid.

(b), (c)$\Rightarrow$ (a). Let $\{x_n\}$ be a sequence in E. Without loss of generality we may assume that it contains infinitely many distinct points, otherwise the conclusion is obvious. Let A denote the set consisting of all points in this sequence. All balls of radius 1 cover X and hence $\overline{A}$. By the finite cover property (see Remark 2.1) there is a finite cover of $\overline{A}$. We can pick one B_1 such that $B_1 \cap A$ contains infinitely many points from A. Next, all balls of radius 1/2 cover $\overline{B_1} \cap \overline{A}$. By the same reasoning, we let $B_{1/2}$ be a ball of radius 1/2 such that $B_{1/2} \cap B_1 \cap A$ contains infinitely many points from A. By repeating this process, we obtain a sequence of balls, $\{B_{1/k}\}$, one of each of radius $1/k$, such that there are infinitely many points of A inside $C_k = B_{1/k} \cap B_{1/(k-1)} \cdots \cap B_1 \cap A$. Pick a point from each C_k to form a subsequence $\{z_k\}$ of $\{x_n\}$. (Here we use the fact that there are infinitely many points of A in C_k to guarantee the existence of the subsequence.) Observing that $\overline{C_k}$ is descending and so enjoys the finite intersection property, $\bigcap_k \overline{C_k}$ is non-empty. Let z be a point in this common intersection. (In fact, there is exactly one point in this common intersection, but we do not need this fact.) As the radius of $B_{1/k}$ tending to 0, $\{z_k\}$ converges to z. We have shown that E is compact. $\square$

In the proof of the following result, we illustrate how to prove the same statement by using the subsequence approach and the finite cover approach.

Proposition 2.5. *Let K be a compact set and G be an open set, $K \subset G$, in the metric space (X, d). Then*

$$\text{dist } (K, \partial G) > 0,$$

where dist $(A, B) = \inf\{d(x, y) : \ x \in A, \ y \in B\}$.

Proof. First proof: Suppose on the contrary that dist $(K, \partial G) = 0$. By the definition of the distance between two sets, there are $\{x_n\} \subset K$ and $\{y_n\} \subset \partial G$ such that $d(x_n, y_n) \to 0$. By the compactness of K, there exists a subsequence $\{x_{n_j}\}$ and $x^* \in K$ such that $x_{n_j} \to x^*$. From $d(x^*, y_{n_j}) \leq d(x_{n_j}, y_{n_j}) + d(x^*, x_{n_j}) \to 0$ we see that $x^* \in \partial G$ (the boundary of a set is always closed). But then $G \cap \partial G$ is non-empty, which is impossible as G is open. So dist $(K, \partial G) > 0$.

Second proof: For $x \in K$, we claim that $d(x, \partial G) > 0$. For, if $d(x, \partial G) = 0$, there exists $\{y_n\} \subset \partial G$, $d(x, y_n) \to 0$, but then x belongs to $\overline{\partial G}$. As the boundary of a set is always a closed set, $x \in \partial G$ contradicting $x \in G$. So $d(x, \partial G) > 0$. Due the continuity of $x \mapsto d(x, \partial G)$, we can find a small number $\rho_x > 0$ such that dist $(y, \partial G) \geq d(x, \partial G)/2 > 0$ for all $y \in B_{\rho_x}(x)$. That is, $\text{dist}(B_{\rho_x}(x), \partial G) \geq d(x, \partial G)/2$. The collection of all balls $B_{\rho_x}(x)$, $x \in K$, forms an open cover of K. Since K is compact, there exist $x_1, \cdots, x_N$, such that $B_{\rho_{x_j}}(x_j)$, $j = 1, \cdots, N$, form a finite subcover of K. Taking $\delta = \min\{d(x_1, \partial G)/2, \cdots, d(x_N, \partial G)/2\}$, we conclude $\text{dist}(K, \partial G) \geq \text{dist}(\cup_j B_{\rho_j}(x_j), \partial G) \geq \delta > 0$. $\square$

We finally note

Proposition 2.6. *Let E be a compact set in (X, d) and $F : (X, d) \to (Y, \rho)$ be continuous. Then $f(E)$ is a compact set in (Y, ρ).*

Proof. Let $\{y_n\}$ be a sequence in $f(E)$ and let $\{x_n\}$ be in E satisfying $f(x_n) = y_n$ for all n. By the compactness of E, there exist some $\{x_{n_j}\}$ and x in E such that $x_{n_j} \to x$ as $j \to \infty$. By the continuity of f, we have $y_{n_j} = f(x_{n_j}) \to f(x)$ in $f(E)$. Hence $f(E)$ is compact. $\square$

Can you prove this property by using the finite cover property of compact sets?

There are several fundamental theorems which hold for continuous functions defined on a closed, bounded set in the Euclidean space. Notably they include

- A continuous function on a closed, bounded set is uniformly continuous; and
- A continuous function on a closed, bounded set attains its minimum and maximum in the set.

Although they may no longer hold on arbitrary closed, bounded sets in a general metric space, they continue to hold when the sets are strengthened to compact ones. The proofs are very much like in the finite dimensional case. I leave them as exercises.

2.6 The Inverse Function Theorem

The Inverse Function Theorem and Implicit Function Theorem play a fundamental role in analysis and geometry. They illustrate the principle of linearization which is ubiquitous in mathematics. We learned these theorems in advanced calculus but the proofs were not emphasized. Now we fill out the gap. Adapting the notations in advanced calculus, a point $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ is sometimes called a vector and we use $|x|$ instead of $\|x\|_2$ to denote its Euclidean norm in this section.

All is about linearization. Recall that a real-valued function on an open interval I is differentiable at some $x_0 \in I$ if there exists some $a \in \mathbb{R}$ such that

$$\lim_{x\to x_0} \left| \frac{f(x) - f(x_0) - a(x-x_0)}{x - x_0} \right| = 0.$$

In fact, the value a is equal to $f'(x_0)$, the derivative of f at x_0. We can rewrite the limit above using the little o notation:

$$f(x_0 + z) - f(x_0) = f'(x_0)z + \circ(z), \quad \text{as } z \to 0.$$

Here $\circ(z)$ denotes a quantity satisfying $\lim_{z\to 0} \circ(z)/|z| = 0$. The same situation carries over to a real-valued function f in some open set in $\mathbb{R}^n$. A function f is called differentiable at x_0 in this open set if there exists a vector $a = (a_1, \cdots, a_n)$ such that

$$f(x_0 + x) - f(x_0) = \sum_{j=1}^{n} a_j x_j + \circ(z) \quad \text{as } x \to 0.$$

Note that here $x_0 = (x_0^1, \cdots, x_0^n)$ is a vector. Again one can show that the vector a is uniquely given by the gradient vector of f at x_0

$$\nabla f(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \cdots, \frac{\partial f}{\partial x_n}(x_0) \right).$$

More generally, a map F from an open set in $\mathbb{R}^n$ to $\mathbb{R}^m$ is called differentiable at a point x_0 in this open set if each component of $F = (f^1, \cdots, f^m)$ is differentiable. We can write the differentiability condition collectively in the following form

$$F(x_0 + x) - F(x_0) = DF(x_0)x + o(x), \tag{2.3}$$

where $DF(x_0)$ is the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ given by

$$(DF(x_0)z)_i = \sum_{j=1}^{n} a_{ij}(x_0)x_j, \quad i = 1, \cdots, m,$$

where $\big(a_{ij}\big) = \big(\partial f^i/\partial x_j\big)$ is the Jabocian matrix of f. (2.3) shows near x_0, that is, when x is small, the function F is well-approximated by the linear map $DF(x_0)$ up to the constant $F(x_0)$ as long as $DF(x_0)$ is nonsingular. It suggests that the local information of a map at a differentiable point could be retrieved from its a linear map, which is much easier to analyse. This principle, called linearization, is widely used in analysis. The Inverse Function Theorem is a typical result of linearization. It asserts that a map is locally invertible if its linearization is invertible. Therefore, local bijectivity of the map is ensured by the invertibility of its linearization. When $DF(x_0)$ is not invertible, the first term on the right hand side of (2.3) may degenerate in some or even all direction so that $DF(x_0)x$ cannot control the error term $\circ(|z|)$. In this case the local behavior of F may be different from its linearization.

Theorem 2.1 (Inverse Function Theorem). *Let $F : U \to \mathbb{R}^n$ be a C^1-map where U is open in $\mathbb{R}^n$ and $x_0 \in U$. Suppose that $DF(x_0)$ is invertible.*

(a) There exist open sets V and W containing x_0 and $F(x_0)$ respectively such that the restriction of F on V is a bijection onto W with a C^1-inverse.

(b) The inverse is C^k when F is $C^k, 1 \le k \le \infty$, in V.

A map from some open set in R^n to $\mathbb{R}^m$ is C^k, $1 \le k \le \infty$, if all its components belong to C^k. It is called a C^∞-map or a smooth map if its components are C^∞. Similarly, a matrix is C^k or smooth if its entries are C^k or smooth accordingly.

The condition that $DF(x_0)$ is invertible, or equivalently the non-vanishing of the determinant of the Jacobian matrix, is called the nondegeneracy condition. Without this condition, the map may or may not be local invertible, see the examples below. Nevertheless, it is necessary for the differentiability of the local inverse. At this point, let us recall the general chain rule.

Let $G : \mathbb{R}^n \to \mathbb{R}^m$ and $F : \mathbb{R}^m \to \mathbb{R}^l$ be C^1 and their composition $H = F \circ G : \mathbb{R}^n \to \mathbb{R}^l$ is also C^1. We compute the first partial derivatives of H in terms of the partial derivatives of F and G. Letting $G = (g_1, \cdots, g_m)$, $F = (f_1, \cdots, f_l)$ and $H = (h_1, \cdots, h_l)$. From

$$h_k(x_1, \cdots, x_n) = f_k(g_1(x), \cdots, g_m(x)), \quad k = 1, \cdots, l,$$

we have

$$\frac{\partial h_k}{\partial y_i} = \sum_{i=1}^n \frac{\partial f_k}{\partial x_i}\frac{\partial g_i}{\partial x_j} \ .$$

Writing it in matrix form we have

$$DF(G(x))DG(x) = DH(x).$$

When the inverse is differentiable, we may apply this chain rule to differentiate the relation $F^{-1}(F(x)) = x$ to obtain

$$DF^{-1}(y_0)\ DF(x_0) = I \ , \quad y_0 = F(x_0),$$

where I is the identity map. We conclude that

$$DF^{-1}(y_0) = \big(DF(x_0)\big)^{-1},$$

in other words, the matrix of the derivative of the inverse map is precisely the inverse matrix of the derivative of the map. We conclude that *in order to have a differentiable inverse near x_0, $DF(x_0)$ must be invertible.*

Lemma 2.1. *Let A be a linear map from $\mathbb{R}^n$ to itself given by*

$$(Ax)_i = \sum_{j=1}^{n} a_{ij}x_j, \quad i = 1, \cdots n.$$

Then

$$|Az| \leq M(A)\ |z|, \quad \forall z \in \mathbb{R}^n,$$

where $M(A) = \sqrt{\sum_{i,j} a_{ij}^2}$.

Proof. By Cauchy-Schwarz inequality,

$$\begin{aligned}
|Ax|^2 &= \sum_i (Ax)_i^2 \\
&= \sum_i \big(\sum_j a_{ij}x_j\big)^2 \\
&\leq \sum_i \big(\sum_j a_{ij}^2\big)\big(\sum_j x_j^2\big) \\
&= M(A)^2\ |x|^2\ .
\end{aligned}$$

□

Now we prove Theorem 2.19. Let $x_0 \in U$ and $y_0 = F(x_0)$. We set $\widetilde{F}(x) = F(x + x_0) - y_0$ in $\widetilde{U} = U - x_0$ so that $\widetilde{F}(0) = 0$ and $D\widetilde{F}(0) = DF(x_0)$. First we would like to show that there is a unique solution for the equation $\widetilde{F}(x) = y$ for y near 0. We will use the Contraction Mapping Principle to achieve our goal. After a further restriction on the size of $\widetilde{U} = U - x_0$, we may assume that $\widetilde{F}$ is C^1 with $D\widetilde{F}(x)$ invertible at all $x \in \widetilde{U}$. For a fixed y, define the map in $\widetilde{U}$ by

$$T(x) = L^{-1}\left(Lx - \widetilde{F}(x) + y\right)$$

where $L = D\widetilde{F}(0)$. It is clear that any fixed point of T is a solution to $\widetilde{F}(x) = y$. By the lemma,

$$\begin{aligned}
|T(x)| &\leqslant M(L^{-1})\ |\widetilde{F}(x) - Lx - y| \\
&\leqslant M(L^{-1})\left(|\widetilde{F}(x) - Lx| + |y|\right) \\
&\leq M(L^{-1})\left(\left|\int_0^1 (D\widetilde{F}(tx) - D\widetilde{F}(0))dt\ x\right| + |y|\right),
\end{aligned}$$

where we have used the formula

$$\widetilde{F}(x) - D\widetilde{F}(0)x = \int_0^1 \frac{d}{dt}\widetilde{F}(tx)dt - D\widetilde{F}(0)x = \int_0^1 \big(D\widetilde{F}(tx) - D\widetilde{F}(0)\big)dt\ x,$$

after using the chain rule to get

$$\frac{d}{dt}\widetilde{F}(tx) = D\widetilde{F}(tx) \cdot x.$$

By the continuity of $D\widetilde{F}$ at 0, we can find a small ρ_0 such that $B_{\rho_0}(0) \subset \widetilde{U}$ and

$$M(L^{-1})M(D\widetilde{F}(x) - D\widetilde{F}(0)) \leq \frac{1}{2}, \quad \forall x \in B_{\rho_0}(0)\ . \tag{2.4}$$

Then for for each y in $B_R(0)$, where R is chosen to satisfy $M(L^{-1})R \leq \rho_0/2$, by (2.4) we have

$$\begin{aligned} |T(x)| &\leq M(L^{-1})\left(\int_0^1 |(D\widetilde{F}(tx) - D\widetilde{F}(0))x|dt + |y|\right) \\ &\leq M(L^{-1})\left(\sup_{t\in[0,1]} M(D\widetilde{F}(tx) - D\widetilde{F}(0))|x| + |y|\right) \\ &\leq \frac{1}{2}|x| + M(L^{-1})|y| \\ &\leq \frac{1}{2}\rho_0 + \frac{1}{2}\rho_0 = \rho_0, \end{aligned}$$

for all $x \in B_{\rho_0}(0)$. We conclude that T maps $\overline{B_{\rho_0}(0)}$ to itself. Moreover, for x_1, x_2 in $B_{\rho_0}(0)$, we have

$$\begin{aligned} |T(x_2) - T(x_1)| &= \left|M(L^{-1})\left(\widetilde{F}(x_2) - Lx_2 - y\right) - L^{-1}\left(\widetilde{F}(x_1) - Lx_1 - y\right)\right| \\ &\leqslant M(L^{-1})\left|\widetilde{F}(x_2) - \widetilde{F}(x_1) - D\widetilde{F}(0)(x_2 - x_1)\right| \\ &\leqslant M(L^{-1})\left|\int_0^1 D\widetilde{F}\left(x_1 + t(x_2 - x_1)\right)(x_2 - x_1)dt - D\widetilde{F}(0)(x_2 - x_1)\right|, \end{aligned}$$

where we have used

$$\begin{aligned} \widetilde{F}(x_2) - \widetilde{F}(x_1) &= \int_0^1 \frac{d}{dt}\widetilde{F}(x_1 + t(x_2 - x_1))dt \\ &= \int_0^1 D\widetilde{F}(x_1 + t(x_2 - x_1))(x_2 - x_1)dt. \end{aligned}$$

Using (2.4) we have,

$$|T(x_2) - T(x_1)| \leq \frac{1}{2}|x_2 - x_1|.$$

We have shown that $T : \overline{B_{\rho_0}(0)} \to \overline{B_{\rho_0}(0)}$ is a contraction. By the Contraction Mapping Principle, there is a unique fixed point for T, in other words, for each y in the ball $B_R(0)$ there is a unique point x in $\overline{B_{\rho_0}(0)}$ solving $\widetilde{F}(x) = y$. Defining $\widetilde{G} : \ B_R(0) \to \overline{B_{\rho_0}(0)} \subset \widetilde{U}$ by setting $\widetilde{G}(y) = x$, G is inverse to $\widetilde{F}$.

Next, we claim that $\widetilde{G}$ is continuous. In fact, for $\widetilde{G}(y_i) = x_i \in B_R(0), \ i = 1, 2$, (not to be mixed up with the x_i above),

$$
\begin{aligned}
|\widetilde{G}(y_2) - \widetilde{G}(y_1)| &= |x_2 - x_1| \\
&= |T(x_2) - T(x_1)| \\
&\leq M(L^{-1}) \left(\left| \widetilde{F}(x_2) - \widetilde{F}(x_1) - L(x_2 - x_1) \right| + |y_2 - y_1| \right) \\
&\leq M(L^{-1}) \left(\left| \int_0^1 \big(D\widetilde{F}((1-t)x_1 + tx_2) - D\widetilde{F}(0) \big) dt (x_2 - x_1) \right| + |y_2 - y_1| \right) \\
&\leq \frac{1}{2} |x_2 - x_1| + M(L^{-1}) |y_2 - y_1| \\
&= \frac{1}{2} |\widetilde{G}(y_2) - \widetilde{G}(y_1)| + M(L^{-1}) |y_2 - y_1|,
\end{aligned}
$$

where (2.4) has been used. We deduce

$$
|\widetilde{G}(y_2) - \widetilde{G}(y_1)| \leqslant 2M(L^{-1}) |y_2 - y_1| \ , \tag{2.5}
$$

that's, $\widetilde{G}$ is Lipschitz continuous on $B_R(0)$.

Finally, let's show that $\widetilde{G}$ is a C^1-map in $B_R(0)$. In fact, for $y_1, y_1 + y$ in $B_R(0)$, using

$$
\begin{aligned}
y &= (y_1 + y) - y_1 \\
&= \widetilde{F}(\widetilde{G}(y_1 + y)) - \widetilde{F}(\widetilde{G}(y_1)) \\
&= \int_0^1 D\widetilde{F}(\widetilde{G}(y_1) + t(\widetilde{G}(y_1 + y) - \widetilde{G}(y_1)) dt \ (\widetilde{G}(y_1 + y) - \widetilde{G}(y_1)),
\end{aligned}
$$

we have

$$
\widetilde{G}(y_1 + y) - \widetilde{G}(y_1) = (D\widetilde{F})^{-1}(\widetilde{G}(y_1))y + R,
$$

where R is given by

$$
(D\widetilde{F})^{-1}(\widetilde{G}(y_1)) \int_0^1 \Big(D\widetilde{F}(\widetilde{G}(y_1)) - D\widetilde{F}(\widetilde{G}(y_1) + t(\widetilde{G}(y_1 + y) - \widetilde{G}(y_1)) \Big) (\widetilde{G}(y_1 + y) - \widetilde{G}(y_1)) dt.
$$

As $\widetilde{G}$ is continuous and $\widetilde{F}$ is C^1, we have

$$
\widetilde{G}(y_1 + y) - \widetilde{G}(y_1) - (D\widetilde{F})^{-1}(\widetilde{G}(y_1))y = \circ(1)(\widetilde{G}(y_1 + y) - \widetilde{G}(y_1))
$$

for small y. Using (2.5), we see that

$$
\widetilde{G}(y_1 + y) - \widetilde{G}(y_1) - (D\widetilde{F})^{-1}(\widetilde{G}(y_1))y = \circ(y) \ ,
$$

as $|y| \to 0$. We conclude that $\widetilde{G}$ is differentiable with Jacobian matrix $(D\widetilde{F})^{-1}(\widetilde{G}(y_1))$.

Going back to the original function, we see that $G(y) = \widetilde{G}(y-y_0)+x_0$ is the C^1-inverse to F from $B_R(y_0)$ back to $\overline{B_{\rho_0}(x_0)} \subset U$.

After proving the differentiability of G, from the formula $DF(G(y))DG(y) = I$ where I is the identity matrix we see that $DG(y) = (DF(G(y))^{-1}$ for $y \in B_R(0)$. From linear algebra we know that each entry of $DG(y)$ can be expressed as a rational function of the entries of the matrix of $DF(G(y))$. Consequently, $DG(y)$ is C^k in y if $DF(G(y))$ is C^k for $1 \leq k \leq \infty$.

The proof of the Inverse Function Theorem is completed by taking $W = B_R(y_0)$ and $V = G(W)$.

Example 2.1. The Inverse Function Theorem asserts a local invertibility. Even if the linearization is non-singular everywhere, we cannot assert global invertibility. Let us consider the switching between the cartesian and polar coordinates in the plane:

$$x = r\cos\theta, \quad y = r\sin\theta\ .$$

The function $F: (0,\infty) \times (-\infty,\infty) \to \mathbb{R}^2$ given by $F(r,\theta) = (x,y)$ is a continuously differentiable function whose Jacobian matrix is non-singular except $(0,0)$. However, it is clear that F is not bijective, for instance, all points $(r, \theta + 2n\pi), n \in \mathbb{Z}$, have the same image under F.

Example 2.2. An exceptional case is dimension one where a global result is available. Indeed, in Mathematical Analysis II we learned that if f is continuously differentiable on (a,b) with non-vanishing f', it is either strictly increasing or decreasing so that its global inverse exists and is again continuously differentiable.

Example 2.3. Consider the map $F: \mathbb{R}^2 \to \mathbb{R}^2$ given by $F(x,y) = (x^2, y)$. Its Jacobian matrix is singular at $(0,0)$. In fact, for any point $(a,b), a > 0$, $F(\pm\sqrt{a}, b) = (a,b)$. We cannot find any open set, no matter how small is, at $(0,0)$ so that F is injective. On the other hand, the map $H(x,y) = (x^3, y)$ is bijective with inverse given by $J(x,y) = (x^{1/3}, y)$. However, as the non-degeneracy condition does not hold at $(0,0)$ so it is not differentiable there. In these cases the Jacobian matrix is singular, so the nondegeneracy condition does not hold. We will see that in order the inverse map to be differentiable, the nondegeneracy condition must hold.

Inverse Function Theorem may be rephrased in the following form.

A C^k-map F between open sets V and W is a "C^k-diffeomorphism" if F^{-1} exists and is also C^k. Let $f_1, f_2, \cdots, f_n$ be C^k-functions defined in some open set in $\mathbb{R}^n$ whose Jacobian matrix of the map $F = (f_1, \cdots, f_n)$ is non-singular at some point x_0 in this open set. By Theorem 4.1 F is a C^k-diffeomorphism between some open sets V and W containing x_0 and $F(x_0)$ respectively. To every function Φ defined in W, there corresponds a function defined in V given by $\Psi(x) = \Phi(F(x))$, and the converse situation holds. Thus every C^k-diffeomorphism gives rise to a "local change of coordinates".

Next we deduce Implicit Function Theorem from Inverse Function Theorem.

Theorem 2.2 (Implicit Function Theorem). *Consider C^1-map $F : U \to \mathbb{R}^m$ where U is an open set in $\mathbb{R}^n \times \mathbb{R}^m$. Suppose that $(x_0, y_0) \in U$ satisfies $F(x_0, y_0) = 0$ and $D_yF(x_0, y_0)$ is invertible in $\mathbb{R}^m$. There exist an open set $V_1 \times V_2$ in U containing (x_0, y_0) and a C^1-map $\varphi : V_1 \to V_2$, $\varphi(x_0) = y_0$, such that*

$$F(x, \varphi(x)) = 0 \ , \qquad \forall x \in V_1 \ .$$

The map φ belongs to C^k when F is $C^k, 1 \leq k \leq \infty$, in U. Moreover, if ψ is another C^1-map in some open set containing x_0 to V_2 satisfying $F(x, \psi(x)) = 0$ and $\psi(x_0) = y_0$, then ψ coincides with φ in their common set of definition.

The notation $D_yF(x_0, y_0)$ stands for the linear map associated to the Jocabian matrix $(\partial F_i/\partial y_j(x_0, y_0))_{i,j=1,\cdots,m}$ where x_0 is fixed. In general, a version of Implicit Function Theorem holds when the rank of DF at a point is m. In this case, we can rearrange the independent variables to make D_yF non-singular at that point.

Proof. Consider $\Phi : U \to \mathbb{R}^n \times R^m$ given by

$$\Phi(x, y) = (x, F(x, y)).$$

It is evident that $D\Phi(x, y)$ is invertible in $\mathbb{R}^n \times \mathbb{R}^m$ when $D_yF(x, y)$ is invertible in $\mathbb{R}^m$. By the Inverse Function Theorem, there exists a C^1-inverse $\Psi = (\Psi_1, \Psi_2)$ from some open W in $\mathbb{R}^n \times \mathbb{R}^m$ containing $\Phi(x_0, y_0)$ to an open subset of U. By restricting W further we may assume $\Psi(W)$ is of the form $V_1 \times V_2$. For every $(x, z) \in W$, we have

$$\Phi(\Psi_1(x, z), \Psi_2(x, z)) = (x, z),$$

which, in view of the definition of Φ, yields

$$\Psi_1(x, z) = x, \text{ and } F((\Psi_1(x, z), \Psi_2(x, z)) = z.$$

In other words, $F(x, \Psi_2(x, z)) = z$ holds. In particular, taking $z = 0$ gives

$$F(x, \Psi_2(x, 0)) = 0, \quad \forall x \in V_1 \ ,$$

so the function $\varphi(x) \equiv \Psi_2(x, 0)$ satisfies our requirement.

By restricting V_1 and V_2 further if necessary, we may assume the matrix

$$\int_0^1 D_yF(x, y_1 + t(y_2 - y_1)dt$$

is nonsingular for $(x, y_1), (x, y_2) \in V_1 \times V_2$. Now, suppose ψ is a C^1-map defined near x_0 satisfying $\psi(x_0) = y_0$ and $F(x, \psi(x)) = 0$. We have

$$\begin{aligned} 0 &= F(x, \psi(x)) - F(x, \varphi(x)) \\ &= \int_0^1 D_y F(x, \varphi(x) + t(\psi(x) - \varphi(x)) dt (\psi(x) - \varphi(x)), \end{aligned}$$

for all x in the common open set they are defined. This identity forces that ψ coincides with φ in this open set. The proof of the implicit function is completed, once we observe that the regularity of φ follows from Inverse Function Theorem. $\square$

Example 2.4. We illustrate the condition det $DF_y(x_0, y_0) \neq 0$ or the equivalent condition

$$\text{rank} DF(p_0) = m, \quad p_0 \in \mathbb{R}^{n+m} \ ,$$

in the following three cases.

First, consider the function $F_1(x, y) = x - y^2 + 3$. We have $F_1(-3, 0) = 0$ and $F_{1x}(-3, 0) = 1 \neq 0$. By Implicit Function Theorem, the zero set of F_1 can be described near $(-3, 0)$ by a function $x = \varphi(y)$ near $y = 0$. Indeed, by solving the equation $F_1(x, y) = 0$, $\varphi(y) = y^2 - 3$. On the other hand, $F_{1y}(-3, 0) = 0$ and from the formula $y = \pm\sqrt{x + 3}$ we see that the zero set is not a graph over an open interval containing -3.

Next we consider the function $F_2(x, y) = x^2 - y^2$ at $(0, 0)$. We have $F_{2x}(0, 0) = F_{2y}(0, 0) = 0$. Indeed, the zero set of F_2 consists of the two straight lines $x = y$ and $x = -y$ intersecting at the origin. It is impossible to express it as the graph of a single function near the origin.

Finally, consider the function $F_3(x, y) = x^2 + y^2$ at $(0, 0)$. We have $F_{3x}(0, 0) = F_{3y}(0, 0) = 0$. Indeed, the zero set of F_3 degenerates into a single point $\{(0, 0)\}$ which cannot be the graph of any function.

It is interesting to note that the Inverse Function Theorem can be deduced from Implicit Function Theorem. Thus they are equivalent. To see this, keeping the notations used in Theorem 2.19. Define a map $\widetilde{F} : U \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$\widetilde{F}(x, y) = F(x) - y.$$

Then $\widetilde{F}(x_0, y_0) = 0, y_0 = F(x_0)$, and $D\widetilde{F}(x_0, y_0)$ is invertible. By Theorem 2.20, there exists a C^1-function φ from near y_0 satisfying $\varphi(y_0) = x_0$ and $\widetilde{F}(\varphi(y), y) = F(\varphi(y)) - y = 0$, hence φ is the local inverse of F.

Chapter 3

The Space of Continuous Functions

3.4 Compactness and Arzela-Ascoli Theorem

We pointed out before that not every closed, bounded set in a metric space is compact. In Section 2.3 a bounded sequence without any convergent subsequence is explicitly displayed to show that a closed, bounded set in $C[a, b]$ needs not be compact. In view of numerous theoretic and practical applications, it is strongly desirable to give a characterization of compact sets in $C[a, b]$. The answer is given by the fundamental Arezela-Ascoli Theorem. This theorem gives a necessary and sufficient condition when a closed and bounded set in $C[a, b]$ is compact. In order to have wider applications, we will work on a more general space $C(K)$, where K is a closed, bounded subset of $\mathbb{R}^n$, instead of $C[a, b]$. Recall that $C(K)$ is a complete, separable space under the sup-norm.

The crux for compactness for continuous functions lies on the notion of equicontinuity. Let X be a subset of $\mathbb{R}^n$. A subset $\mathcal{F}$ of $C(X)$ is **equicontinuous** if for every $\varepsilon > 0$, there exists some δ such that

$$|f(x) - f(y)| < \varepsilon, \quad \text{for all } f \in \mathcal{F}, \quad \text{and} \quad |x - y| < \delta, \ x, y \in X.$$

Recall that a function is uniformly continuous in X if for each $\varepsilon > 0$, there exists some δ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta, \ x, y \in X$. So, equicontinuity means that δ can further be chosen independent of the functions in $\mathcal{F}$.

There are various ways to show that a family of functions is equicontinuous. Recall that a function f defined in a subset X of $\mathbb{R}^n$ is called Hölder continuous if there exists some $\alpha \in (0, 1)$ such that

$$|f(x) - f(y)| \le L|x - y|^\alpha, \quad \text{for all } x, y \in X, \tag{3.1}$$

for some constant L. The number α is called the Hölder exponent. The function is called Lipschitz continuous if (3.1) holds for α equals to 1. A family of functions $\mathcal{F}$ in $C(X)$

is said to satisfy a uniform Hölder or Lipschitz condition if all members in $\mathcal{F}$ are Hölder continuous with the same α and L or Lipschitz continuous and (3.1) holds for the same constant L. Clearly, such $\mathcal{F}$ is equicontinuous. In fact, for any $\varepsilon > 0$, any δ satisfying $L\delta^\alpha < \varepsilon$ can do the job. The following situation is commonly encountered in the study of differential equations. The philosophy is that equicontinuity can be obtained if there is a good, uniform control on the derivatives of functions in $\mathcal{F}$.

Proposition 3.1. *Let $\mathcal{F}$ be a subset of $C(X)$ where X is a convex set in $\mathbb{R}^n$. Suppose that each function in $\mathcal{F}$ is differentiable and there is a uniform bound on the partial derivatives of these functions in $\mathcal{F}$. Then $\mathcal{F}$ is equicontinuous.*

Proof. For, x and y in X, $(1-t)x+ty$, $t \in [0,1]$, belongs to X by convexity. Let $\psi(t) \equiv f((1-t)x+ty)$. By the chain rule

$$\psi'(t) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}((1-t)x+ty)(y_j - x_j),$$

we have

$$\begin{aligned} f(y) - f(x) &= \psi(1) - \psi(0) \\ &= \int_0^1 \psi'(t)dt \\ &= \sum_{j=1}^n \int_0^1 \frac{\partial f}{\partial x_j}(x + t(y-x))(y_j - x_j). \end{aligned}$$

Therefore,

$$|f(y) - f(x)| \leq \sqrt{n} M |y - x|,$$

where $M = \sup\{|\partial f/\partial x_j(x)| : \ x \in X, \ j = 1, \ldots, n, \ f \in \mathcal{F}\}$ after using Cauchy-Schwarz inequality. We conclude that $\mathcal{F}$ satisfies a uniform Lipschitz condition with Lipschitz constant $n^{1/2}M$.

□

Example 3.1. Let

$$\mathcal{A} = \{x : \ x' = \sin(tx), \ t \in [-1,1]\} \subset C[-1,1].$$

It can be shown that, given any $x_0 \in \mathbb{R}$, there is a unique solution x solving the equation and $x(0) = x_0$, so $\mathcal{A}$ contains many functions. There is an obvious uniform estimate on its derivative, namely,

$$|x'(t)| = |\sin tx| \leq 1.$$

By Proposition 3.7 A forms an equicontinuous family. However, as there is no control on x_0, $\mathcal{A}$ is not bounded.

Example 3.2. Let

$$\mathcal{B} = \{f \in C[0,1] : |f(x)| \leq 1, x \in [0,1]\} \subset C[0,1].$$

Clearly $\mathcal{B}$ is closed and bounded. However, we do not have any uniform control on the oscillation of the functions in this set, so it should not be equicontinuous. In fact, consider the sequence $\{\sin nx\}, n \geq 1$, in $\mathcal{B}$. We claim that it is not equicontinuous. In fact, suppose for $\varepsilon = 1/2$, there exists some δ such that $|\sin nx - \sin ny| < 1/2$, whenever $|x - y| < \delta$ for all n. Pick a large n such that $n\delta > \pi$. Taking $x = 0$ and $y = \pi/2n$, $|x - y| < \delta$ but $|\sin nx - \sin ny| = |\sin \pi/2| = 1 > 1/2$, contradiction holds. Hence $\mathcal{B}$ is not equicontinuous.

More examples of equicontinuous families can be found in the exercise.

We first establish a necessary condition for compactness.

Theorem 3.2 (**Arzela's Theorem**). *A compact set in $C(K)$ where K is a compact set in $\mathbb{R}^n$ is closed, bounded and equicontinuous.*

Proof. In general, every compact set in a metric space is closed and bounded. Let $\mathcal{F}$ be compact in $C(K)$. It remains to prove equicontinuity. Since a compact set is totally bounded, for each $\varepsilon > 0$, there exist $f_1, \cdots, f_N \in \mathcal{F}$ such that $\mathcal{F} \subset \bigcup_{j=1}^{N} B_\varepsilon(f_j)$ where N depends on ε. So for any $f \in \mathcal{F}$, there exists f_j such that

$$|f(x) - f_j(x)| < \varepsilon, \quad \text{for all } x \in K.$$

As each f_j is continuous, there exists δ_j such that $|f_j(x) - f_j(y)| < \varepsilon$ whenever $|x - y| < \delta_j$. Letting $\delta = \min\{\delta_1, \cdots, \delta_N\}$, then

$$|f(x) - f(y)| \leq |f(x) - f_j(x)| + |f_j(x) - f_j(y)| + |f_j(y) - f(y)| < 3\varepsilon,$$

for $|x - y| < \delta$, so $\mathcal{F}$ is equicontinuous.

□

It turns out the converse of Arzela's theorem is also true.

Theorem 3.3 (**Ascoli's Theorem**). *A closed, bounded and equicontinuous set in $C(K)$ where K is a compact set in $\mathbb{R}^n$ is compact.*

We need the following useful lemma from elementary analysis.

Lemma 3.4. *Let A be a countable set and $\{f_n\}$ be a sequence of real-valued functions defined on A. Suppose that for each $z \in A$, there exists an M such that $|f_n(z)| \leq M$ for all $n \geq 1$. There is a subsequence of $\{f_n\}$, $\{f_{n_k}\}$, such that $\{f_{n_k}(z)\}$ is convergent at each $z \in A$.*

Proof. Let $A = \{z_j\}, j \geq 1$. Since $\{f_n(z_1)\}$ is a bounded sequence, we can extract a subsequence $\{f_n^1\}$ such that $\{f_n^1(z_1)\}$ is convergent. Next, as $\{f_n^1(z_2)\}$ is bounded, it has a subsequence $\{f_n^2\}$ such that $\{f_n^2(z_2)\}$ is convergent. Keep doing in this way, we obtain sequences $\{f_n^j\}$ satisfying (i) $\{f_n^{j+1}\}$ is a subsequence of $\{f_n^j\}$ and (ii) $\{f_n^j(z_1)\}, \{f_n^j(z_2)\}, \cdots, \{f_n^j(z_j)\}$ are convergent. Then the diagonal sequence $\{g_n\},\ g_n = f_n^n,$ for all $n \geq 1$, is a subsequence of $\{f_n\}$ which converges at every z_j. □

The subsequence selected in this way is usually called a Cantor's diagonal sequence. It first came up in Cantor's study of infinite sets.

Proof of Ascoli's Theorem. Since K is compact in $\mathbb{R}^n$, it is totally bounded. For each $j \geq 1$, we can cover K by finitely many balls $B_{1/j}(x_1^j), \cdots, B_{1/j}(x_N^j)$ where the number N depends on j. All $\{x_k^j\},\ j \geq 1, 1 \leq k \leq N$, form a countable set. For any sequence $\{f_n\}$ in $\mathcal{F}$, by Lemma 3.10, we can pick a subsequence denoted by $\{g_n\}$ such that $\{g_n(x_k^j)\}$ is convergent for all x_k^j's. We claim that $\{g_n\}$ is a Cauchy sequence in $C(K)$. For, due to the equicontinuity of $\mathcal{F}$, for every $\varepsilon > 0$, there exists a δ such that $|g_n(x) - g_n(y)| < \varepsilon$, whenever $|x - y| < \delta$. Pick $j_0,\ 1/j_0 < \delta$. Then for $x \in K$, there exists $x_k^{j_0}$ such that $|x - x_k^{j_0}| < 1/j_0 < \delta$,

$$\begin{aligned} |g_n(x) - g_m(x)| &\leq |g_n(x) - g_n(x_k^{j_0})| + |g_n(x_k^{j_0}) - g_m(x_k^{j_0})| + |g_m(x_k^{j_0}) - g_m(x)| \\ &< \varepsilon + |g_n(x_k^{j_0}) - g_m(x_k^{j_0})| + \varepsilon. \end{aligned}$$

As $\{g_n(x_k^{j_0})\}$ converges, there exists n_0 such that

$$|g_n(x_k^{j_0}) - g_m(x_k^{j_0})| < \varepsilon, \quad \text{for all } n, m \geq n_0. \tag{3.2}$$

Here n_0 depends on $x_k^{j_0}$. As there are finitely many $x_k^{j_0}$'s, we can choose some N_0 such that (3.2) holds for all $x_k^{j_0}$ and $n, m \geq N_0$. It follows that

$$|g_n(x) - g_m(x)| < 3\varepsilon, \quad \text{for all } n, m \geq N_0,$$

i.e., $\{g_n\}$ is a Cauchy sequence in $C(K)$. By the completeness of $C(K)$ and the closedness of $\mathcal{F}$, $\{g_n\}$ converges to some function in $\mathcal{F}$.

These two theorems together form a necessary and sufficient condition for compactness in $C(K)$. When it comes to applications, the sufficient condition is more relevant than the necessary one. Ascoli's theorem is usually used in the following form.

Theorem 3.5. *A sequence in $C(K)$ where K is compact in $\mathbb{R}^n$ has a convergent subsequence if it is uniformly bounded and equicontinuous.*

Proof. Let $\mathcal{F}$ be the closure of the sequence $\{f_n\}$. We would like to show that $\mathcal{F}$ is bounded and equicontinuous. First of all, by the uniform boundedness assumption, there is some M such that

$$|f_n(x)| \leq M, \quad \forall x \in K,\ n \geq 1.$$

As every function in $\mathcal{F}$ is either one of these f_n or the limit of its subsequence, it also satisfies this estimate, so $\mathcal{F}$ is bounded in $C(K)$. On the other hand, by equicontinuity, for every $\varepsilon > 0$, there exists some δ such that

$$|f_n(x) - f_n(y)| < \frac{\varepsilon}{2}, \quad \forall x, y \in K, \ |x - y| < \delta.$$

As every $f \in \mathcal{F}$ is the limit of a subsequence of $\{f_n\}$, f satisfies

$$|f(x) - f(y)| \leq \frac{\varepsilon}{2} < \varepsilon, \quad \forall x, y \in K, \ |x - y| < \delta,$$

so $\mathcal{F}$ is also equicontinuous. Now the conclusion follows from Ascoli's Theorem.

□

We present an application of Ascoli's Theorem to ordinary differential equations. Consider the initial value problem again,

$$\begin{cases} \dfrac{dx}{dt} = f(t, x), \\ \\ x(t_0) = x_0. \end{cases} \tag{IVP}$$

where f is a continuous function defined in the rectangle $R = [t_0 - a, t_0 + a] \times [x_0 - b, x_0 + b]$. In Chapter 2 we proved that this Cauchy problem has a *unique* solution when f satisfies the Lipschitz condition. Now we show that the existence part of Picard-Lindelöf Theoorem is still valid without the Lipschitiz condition.

We start with an improvement on the domain of existence for the unique solution in Picard-Lindelöf Theorem. Recall that it was shown the solution exists on the interval $(t_0 - a_1, t_0 + a_1)$ where

$$a_1 = \min\left\{a, \frac{b}{M}, \frac{1}{L}\right\} .$$

Now we have

Proposition 3.6. *Under the setting of Picard-Lindelöf Theorem, the unique solution exists on the interval* $(t_0 - a^*, t_0 + a^*)$ *where*

$$a^* = \min\left\{a, \frac{b}{M}\right\} .$$

Proof. Let $w_+(t) = M(t - t_0) + x_0$ and $w_-(t) = -M(t - t_0) + x_0$ where $M = \sup_R |f|$ as before. From $|x'(t)| \leq M$ and $x(t_0) = x_0$ it is easy to see that the solution of (IVP) satisfies $w_-(t) \leq x(t) \leq w_+(t)$ as long as it exists. Let us assume $b/M \leq a$ so that $a^* = b/M$ (the other case can be handled in the same way). The straight lines $x = M(t - t_0) + x_0$ and

$x = -M(t - t_0) + x_0, t \geq t_0$, intersects $x = x_0 + b$ and $x = x_0 - b$ respectively at two points P and Q. The graph of $x(t)$ stays inside the triangle with vertices at $(t_0, x_0), P$ and Q for $t \geq t_0$. We are going to show it exists on $[t_0, t_0 + a^*)$. Let

$$\alpha = \sup\{a: \text{ the solution exists on } (t_0, t_0 + a)\} > 0 \ .$$

Assume on the contrary $\alpha < a^*$. The vertical line $x = \alpha$ intersects $x = M(t - t_0) + x_0$ and $x = -M(t-t_0)+x_0$ respectively at P' and Q'. The triangle Δ with vertices at $(t_0, x_0), P'$ and Q' is a compact set contained in the interior of R. Therefore, $\rho = d(\Delta, \partial R) > 0$. We can find a square $S = [-s, s] \times [-s, s], s = \rho/2$, so that $S_{x,t} \equiv S + (t, x)$ is contained inside R for every $(t, x) \in \Delta$. By applying Picard-Lindelöf Theorem to the equation with initial point (t, x) we conclude that there is a unique solution of the equation over some interval with center at t whose length l is independent of (x, t). Now, since our solution $x(t)$ exists on every closed subinterval of $[0, \alpha)$, we can fix some $t_1, \alpha - t_1 < l/2$. We solve the equation using $(t_1, x(t_1))$ as our initial point to get a solution x_1 defined on an interval of length at least l around t_1. But, using the uniqueness property, the solution $x(t)$ and $x_1(t)$ together formed a solution whose domain extends beyond α, contradiction holds. We conclude that the solution exists up to $t_0 + a^*$. Similarly, we could show that the solution exists on $(t_0 - a^*, t_0]$.

□

Theorem 3.7 (**Cauchy-Peano Theorem**). *Consider (IVP) where f is continuous on $R = [t_0 - a, t_0 + a] \times [x_0 - b, x_0 + b]$. There exist $a' \in (0, a)$ and a C^1-function $x : [t_0 - a', t_0 + a'] \to [x_0 - b, x_0 + b]$, solving (IVP).*

From the proof we will see that a' can be taken to be any number in $(0, \min\{a, b/M\})$ where $M = \sup\{|f(t, x)| : (t, x) \in R\}$. The theorem is also valid for systems.

Proof. By Weierstrass Approximation Theorem, there exists a sequence of polynomials $\{p_n\}$ approaching f in $C(R)$ uniformly. In particular, we have $M_n \to M$, where

$$M_n = \max\{|p_n(t, x)| : (t, x) \in R\}.$$

As each p_n satisfies the Lipschitz condition according to Proposition 3.7, there is a unique solution x_n defined on $I_n = (t_0 - a_n, t_0 + a_n), a_n = \min\{a, b/M_n\}$ for the initial value problem

$$\frac{dx}{dt} = p_n(t, x), \quad x_0(t_0) = x_0.$$

(The Lipschitz constants may depend on p_n, but this does no harm.) As $a_n \to a^* \equiv \min\{a, b/M\}$, the domain of existence I_n eventually expands to $(t_0 - a^*, t_0 + a^*)$ as $n \to \infty$.

Let $a' < a^*$ be fixed. There exists some n_0 such that x_n is well-defined on $[t_0 - a', t_0 + a']$ for all $n \geq n_0$. From $|dx_n/dt| \leq M_n$ and $\lim_{n\to\infty} M_n = M$, we can fix some $n_1 \geq n_0$ such that $M_n \leq M+1$ for all $n \geq n_1$. It follows from Proposition 3.7 that $\{x_n\}$ forms an equicontinuous family on $[t_0 - a', t_0 + a']$. On the other hand, from

$$x_n(t) = x_0 + \int_{t_0}^{t} p_n(s, x_n(s))ds, \tag{3.3}$$

we have

$$|x_n(t)| \leq |x_0| + aM_n \leq |x_0| + a(M+1), \quad n \geq n_1.$$

It follows that $\{x_n\}, n \geq n_1$, is bounded on $[t_0 - a', t_0 + a']$. By Theorem 3.11, it contains a subsequence $\{x_{n_j}\}$ converging uniformly to a continuous function x^* on $[t_0 - a', t_0 + a']$. It remains to check that x^* solves (IVP) on this interval. Passing limit in (3.3), clearly its left hand side tends to $x^*(t)$. To treat its right hand side, we observe that, for $\varepsilon > 0$, there exists some δ such that

$$|f(s_2, x_2) - f(s_1, x_1)| < \varepsilon, \quad |s_2 - s_1|, |x_2 - x_1| < \delta \ . \tag{3.4}$$

This is because f is uniformly continuous on R. Next, there is some $n_2 \geq n_1$ such that $\|f - p_n\|_\infty < \varepsilon$, for all $n \geq n_2$ on R. It implies

$$|p_n(s, x) - f(s, x)| < \varepsilon, \quad \forall (s, x) \in R \ . \tag{3.5}$$

Putting (3.4) and (3.5) together, we have

$$\begin{aligned}
& \left| \int_{t_0}^{t} p_{n_j}(s, x_{n_j}(s))ds - \int_{t_0}^{t} f(s, x^*(s))ds \right| \\
\leq \ & \left| \int_{t_0}^{t} |p_{n_j}(s, x_{n_j}(s)) - f(s, x_{n_j}(s))|ds \right| + \left| \int_{t_0}^{t} |f(s, x_{n_j}(s)) - f(s, x^*(s))|ds \right| \\
\leq \ & 2a\varepsilon \ ,
\end{aligned}$$

which implies the right hand side of (3.3) tends to

$$x_0 + \int_{t_0}^{t} f(s, x^*(s))ds \ ,$$

as $n_j \to \infty$. We conclude that x^* is a solution to (IVP) on $[t_0 - a', t_0 + a']$. □

Chapter 3

The Space of Continuous Functions

3.5 Completeness and Baire Category Theorem

In this section we discuss Baire category theorem, a basic property of complete metric spaces. It is concerned with the decomposition of a metric space into a countable union of subsets. Let us start by looking at the size of a set. Consider the following sets in $\mathbb{R}$ under the usual Euclidean metric:

$$\mathbb{R}, \quad \mathbb{R} \setminus \{a_1, \cdots, a_n\}, \quad \mathbb{I}, \quad \mathbb{Q}\ .$$

Although all of them are dense, their "size" are quite different. You may agree that the first two sets, which are open and dense, are "denser" than the other two. However, $\mathbb{I}$ and $\mathbb{Q}$ are still quite different. How can we distinguish them? As we will see, the situation is better when working on a complete metric space.

Theorem 3.14 (**Baire Category Theorem**). *Let (X, d) be a complete metric and $\{G_n\}$ be a sequence of open, dense sets in X. Then the set $E = \bigcap_{n=1}^{\infty} G_n$ is dense.*

Order $\mathbb{Q}$ as a sequence $\{q_n\}$ and set $G_n = \mathbb{R} \setminus \{q_n\}$. It follows from this theorem that $\mathbb{I} = \bigcap_{n=1}^{\infty} G_n$ is dense. Of course, this is a well-known fact and we re-derive it just as an illustration of the theorem. Note that in the assumption we need the underlying space to be complete. Without this assumption the assertion of the theorem may not hold. Here is an example. Consider the metric space $(\mathbb{Q}, d_2)$ where d_2 is the Euclidean metric. It is routine to check that each $\mathbb{Q} \setminus \{q_n\}$ is open and dense in $\mathbb{Q}$, and yet $\bigcap_{n=1}^{\infty} \mathbb{Q} \setminus \{q_n\}$ is the empty set, which cannot be dense in $\mathbb{Q}$. Here Baire Category Theorem is not applicable because $\mathbb{Q}$ is not complete in the Euclidean metric. Think of the sequence $\{3, 3.1, 3.14, 3, 141, 3.1415, 3.14159, \cdots\}$ which is a Cauchy sequence without limit in $\mathbb{Q}$.

We observe that a set is dense if and only if its complement has empty interior. Indeed, let E be dense in X. Each metric ball must intersect E, thus no ball can be contained in its complement. Its complement must have empty interior. Conversely, if the complement of E has empty interior, it cannot contain any ball. In other words, any ball must intersect E, so E is dense. Using this observation, Baire Category Theorem can be formulated in the following form.

Theorem 3.15 (Baire Category Theorem). *Let (X, d) be a complete metric and $\{F_n\}_1^\infty$ be a sequence of closed set with empty interior. Then the set $\bigcup_{n=1}^\infty F_n$ has empty interior.*

Proof. Let B_0 be any ball. The theorem will be established if we can show that $B_0 \bigcap (X \setminus \bigcup_n F_n) \neq \phi$. As F_1 has empty interior, there exists some point $x_1 \in B_0$ lying outside F_1. Since F_1 is closed , we can find a closed ball $\overline{B}_1 \subset B_0$ centering at x_1 such that $\overline{B}_1 \cap F_1 = \phi$ and its diameter $d_1 \leq d_0/2$, where d_0 is the diameter of B_0. Next, as F_2 has empty interior and closed, by the same reason there is a closed ball $\overline{B}_2 \subset B_1$ centering at x_2 such that $\overline{B}_2 \cap F_2 = \phi$ and $d_2 \leq d_1/2$. Repeating this process, we obtain a sequence of closed balls $\overline{B}_n$ with center x_n satisfying (a) $\overline{B}_{n+1} \subset B_n$, (b) $d_n \leq d_0/2^n$, and (c) $\overline{B}_n$ is disjoint from $F_1, \cdots, F_n$. As the balls shrink to a point, $\{x_n\}$ is a Cauchy sequence. By the completeness of X, $\{x_n\}$ converges to some x^*. As each $\overline{B_n}$ is closed and $x_j \in \overline{B_n}$ for all $j \geq n$, $x^* \in \overline{B_n}$ for all n. In particular, it means that x^* does not belong to F_n for all n, so x^* is a point in B_0 not in $\bigcup_n F_n$.

□

Corollary 3.16. *It is impossible to find closed sets with empty interior, $\{F_n\}, n \geq 1$, in a complete metric space X, so that*

$$X = \bigcup_{n=1}^{\infty} F_n \ .$$

In other, whenever we have decompostion

$$X = \bigcup_{n=1}^{\infty} E_n \ ,$$

for some sequence of closed sets $\{E_n\}$, at least one of these E_n's has non-empty interior.

Proof. By Baire Category Theorem, $\bigcup_{n=1}^\infty F_n$ has empty interior. On the other hand, X is the entire space and $X^o = X$ is non-empty. Therefore, $X = \bigcup_n F_n$ is impossible.

□

Baire category theorem enables us to make a more precise description of the size of a set in a complete metric space. A set in a metric space is called **of second category** if it is the countable intersection of open, dense sets or it contains such a set. It is **of first category** if it is the countable union of closed sets with empty interior, or it is a subset of such a set. From the definition we see that a set is of first category if and only if its complement is of second category. According to Baire Category Theorem, a set of second category is a dense set when the underlying space is complete. However, a set of first category may or may not be dense. Returning to our example in the opening of this section, the sets $\mathbb{R}, \mathbb{R} \setminus \{a_1, \cdots, a_n\}$ and $\mathbb{I}$ are of second category but $\mathbb{Q}$ is of first category. Every finite set in $\mathbb{R}$ is of first category, so is every countable set.

Proposition 3.17. *If a set in a complete metric space is of first category, it cannot be of second category, and vice versa.*

Proof. Let E be of first category and let $E \subset \bigcup_{n=1}^\infty F_n$ where F_n's are closed with empty interior. If it is also of second category, its complement is of first category. Thus, $X \setminus E = \bigcup_{n=1}^\infty E_n$ where E_n's are closed with empty interior. We put F_n's and E_n's together to form a sequence $\{H_n\} = \{F_1, E_1, F_2, E_2, \cdots\}$. Then

$$X = E \bigcup (X \setminus E) \subset \bigcup_{n=1}^\infty H_n ,$$

which contradicts the corollary above. □

Baire Category Theorem has many interesting applications. We end this section by giving two standard ones. It is concerned with the existence of continuous, but nowhere differentiable functions. We knew that Weierstrass is the first person who constructed such a function in 1896. His function is explicitly given in the form of an infinite series

$$W(x) = \sum_{n=1}^\infty \frac{\cos 3^n x}{2^n}.$$

Here we use an implicit argument to show there are far more such functions than continuously differentiable functions.

We begin with a lemma.

Lemma 3.18. *Let $f \in C[a,b]$ be differentiable at x. Then it is Lipschitz continuous at x.*

Proof. By differentiability, for $\varepsilon = 1$, there exists some δ_0 such that

$$\left| \frac{f(y) - f(x)}{y - x} - f'(x) \right| < 1, \quad \forall y \neq x, |y - x| < \delta_0.$$

We have

$$|f(y) - f(x)| \leq L|y - x|, \quad \forall y, |y - x| < \delta_0,$$

where $L = |f'(x)| + 1$. For y lying outside $(x - \delta_0, x + \delta_0)$, $|y - x| \geq \delta_0$. Hence

$$\begin{aligned} |f(y) - f(x)| &\leq |f(x)| + |f(y)| \\ &\leq \frac{2M}{\delta_0}|y - x|, \quad \forall y \in [a, b] \setminus (x - \delta_0, x + \delta_0), \end{aligned}$$

where $M = \sup\{|f(x)| : x \in [a, b]\}$. It follows that f is Lipschitz continuous at x with an Lipschitz constant not exceeding $\max\{L, 2M/\delta_0\}$.

□

Theorem 3.19. *The set of all continuous, nowhere differentiable functions forms a set of second category in $C[a, b]$ and hence dense in $C[a, b]$.*

Proof. For each $L > 0$, define

$S_L = \big\{f \in C[a, b] : f$ is Lipschitz continuous at some x with the Lipschitz constant $\leq L\big\}$.

We claim that S_L is a closed set. For, let $\{f_n\}$ be a sequence in S_L which is Lipschitz continuous at x_n and converges uniformly to f. We need to show $f \in S_L$. By passing to a subsequence if necessary, we may assume $\{x_n\}$ to some x^* in $[a, b]$. We have, by letting $n \to \infty$,

$$\begin{aligned} |f(y) - f(x^*)| &\leq |f(y) - f_n(y)| + |f_n(y) - f(x^*)| \\ &\leq |f(y) - f_n(y)| + |f_n(y) - f_n(x_n)| + |f_n(x_n) - f_n(x^*)| + |f_n(x^*) - f(x^*)| \\ &\leq |f(y) - f_n(y)| + L|y - x_n| + L|x_n - x^*| + |f_n(x^*) - f(x^*)| \\ &\to L|y - x^*| \end{aligned}$$

Next we show that each S_L is nowhere dense. Let $f \in S_L$. By Weierstrass approximation theorem, for every $\varepsilon > 0$, we can find some polynomial p such that $\|f - p\|_\infty < \varepsilon/2$. Since every polynomial is Lipschitz continuous, let the Lipschitz constant of p be L_1. Consider the function $g(x) = p(x) + (\varepsilon/2)\varphi(x)$ where φ is the jig-saw function of period $2r$ satisfying $0 \leq \varphi \leq 1$ and $\varphi(0) = 1$. The slope of this function is either $1/r$ or $-1/r$. Both will become large when r is chosen to be small. Clearly, we have $\|f - g\|_\infty < \varepsilon$. On the other hand,

$$\begin{aligned} |g(y) - g(x)| &\geq \frac{\varepsilon}{2}|\varphi(y) - \varphi(x)| - |p(y) - p(x)| \\ &\geq \frac{\varepsilon}{2}|\varphi(y) - \varphi(x)| - L_1|y - x| \ . \end{aligned}$$

For each x sitting in $[a, b]$, we can always find some y nearby so that the slope of φ over the line segment between x and y is greater than $1/r$ or less than $-1/r$. Therefore, if we choose r so that

$$\frac{\varepsilon}{2}\frac{1}{r} - L_1 > L,$$

we have $|g(y) - g(x)| > L|y - x|$, that is, g does not belong to S_L.

Denoting by S the set of functions in $C[a,b]$ which are differentiable at at least one point, by the lemma it must belong to S_N for some positive integer N. Therefore, $S \subset \cup_{N=1}^{\infty} S_N$ is of first category. Its complement, the collection of continuous, nowhere differentiable functions, is of second category, and hence dense in $C[0,1]$. The completeness of $C[0,1]$ is always in effect. $\square$

Though elegant, a drawback of this proof is that one cannot assert which particular function is nowhere differentiable. On the other hand, the example of Weierstrass is a concrete one.

Our second application is concerned with the basis of a vector space. Recall that a basis of a vector space is a set of linearly independent vectors such that every vector can be expressed as a linear combination of vectors from the basis. The construction of a basis in a finite dimensional vector space was done in MATH2040. However, in an infinite dimensional vector space the construction of a basis is difficult. Nevertheless, using Zorn's lemma, a variant of the axiom of choice, one shows that a basis always exists. Some authors call a basis for an infinite dimensional basis a Hamel basis. The difficulty in writing down a Hamel basis is explained in the following result.

Theorem 3.20. *Every basis of an infinite dimensional Banach space consists of uncountably many vectors.*

Proof. Let V be an infinite dimensional Banach space and $\mathcal{B} = \{w_j\}$ be a countable basis. We aim for a contradiction. Indeed, let W_n be the subspace spanned by $\{w_1, \cdots, w_n\}$. We have the decomposition

$$V = \bigcup_n W_n .$$

If one can show that each W_n is closed and has empty interior, since V is complete, Corollary 3.16 tells us this decomposition is impossible. To see that W_n has empty interior, pick a unit vector v_0 outside W_n. For $w \in W_n$ and $\varepsilon > 0$, $w + \varepsilon v_0 \in B_\varepsilon(w) \cap (V \setminus W_n)$, so W_n cannot contain a ball. Next, letting v_j be a sequence in W_n and $v_j \to v_0$, we would like to show that $v \in W_n$. Indeed, every vector $v \in W_n$ can be uniquely expressed as $\sum_{j=1}^n a_j w_j$. The map $v \mapsto a \equiv (a_1, \cdots, a_n)$ sets up a linear bijection between W_n and $\mathbb{R}^n$ and $|||a||| \equiv \|v\|$ defines a norm on $\mathbb{R}^n$. Since any two norms in $\mathbb{R}^n$ are equivalent (see exercise), a convergent (resp. Cauchy) sequence in one norm is the same in the other norm. Since now $\{v_j\}$ is convergent in V, it is a Cauchy sequence in V. The corresponding sequence $\{a^j\}$, $a^j = (a_1^j, \cdots, a_n^j)$, is a Cauchy sequence in $\mathbb{R}^n$ with respect to $|||\cdot|||$ and hence in $\|\cdot\|_2$, the Euclidean norm. Using the completeness of $\mathbb{R}^n$ with respect to the Euclidean norm, $\{a^j\}$ converges to some $a^* = (a_1^*, \cdots, a_n^*)$. But then $\{v_j\}$ converges to $v^* = \sum_j a_j^* w_j$ in W_n. By the uniqueness of limit, we conclude that $v_0 = v^* \in W_n$, so W_n is closed. $\square$

In some books, the concept of a nowhere dense set is introduced. Indeed, a set in a metric space is nowhere dense if it is contained in a closed set with empty interior. Baire Category Theorem can be rephrased as, the countable intersection of nowhere dense sets has empty interior. I avoid it for possible confusion. You already have names like dense, empty interior, first category and second category sets, and that is enough!

Comments on Chapter 3. Three properties, namely, separability, compactness, and completeness of the space of continuous functions are studied in this chapter.

Separability is established by various approximation theorems. For the space $C[a,b]$, Weierstrass approximation theorem is applied. Weierstrass (1885) proved his approximation theorem based on the heat kernel, that is, replacing the kernel Q_n in our proof in Section 1 by the heat kernel. The argument is a bit more complicated but essentially the same. It is taken from Rudin, Principles of Mathematical Analysis. A proof by Fourier series is already presented in Chapter 1. Another standard proof is due to Bernstein, which is constructive and had initiated a branch of analysis called approximation theory. The Stone-Weierstrass theorem is due to M.H. Stone (1937, 1948). We use it to establish the separability of the space $C(X)$ where X is a compact metric space. You can find more approximation theorem by web-surfing.

Arzela-Ascoli theorem plays the role in the space of continuous functions the same as Bolzano-Weierstrass theorem does in the Euclidean space. A bounded sequence of real numbers always admits a convergent subsequence. Although this is no longer true for bounded sequences of continuous functions on $[a,b]$, it does hold when the sequence is also equicontinuous. Ascoli's theorem (Proposition 3.11) is widely applied in the theory of partial differential equations, the calculus of variations, complex analysis and differential geometry. Here is a taste of how it works for a minimization problem. Consider

$$\inf \left\{ J[u] : \ u(0) = 0, u(1) = 5, \ u \in C^1[0,1] \right\},$$

where

$$J[u] = \int_0^1 \left(u'^2(x) - \cos u(x) \right) dx.$$

First of all, we observe that $J[u] \geq -1$. This is clear, for the cosine function is always bounded by ± 1. After knowing that this problem is bounded from -1, we see that $\inf J[u]$ must be a finite number, say, γ. Next we pick a minimizing sequence $\{u_n\}$, that is, every u_n is in $C^1[0,1]$ and satisfies $u(0) = 0, u(1) = 5$, such that $J[u_n] \to \gamma$ as $n \to \infty$. By

Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\big|u_n(x)-u_n(y)\big| &\le \int_x^y \big|u_n'(x)\big|dx \\
&\le \sqrt{\int_x^y 1^2dx}\sqrt{\int_x^y u_n'^2(x)dx} \\
&\le \sqrt{\int_x^y 1^2dx}\sqrt{\int_0^1 u_n'^2(x)dx} \\
&\le \sqrt{J[u_n]+1}\sqrt{|y-x|} \\
&\le \sqrt{\gamma+2}\ |y-x|^{1/2}
\end{aligned}$$

for all large n. From this estimate we immediately see that $\{u_n\}$ is equicontinuous and bounded (because $u_n(0)=0$). By Ascoli's theorem, it has a subsequence $\{u_{n_j}\}$ converging to some $u \in C[0,1]$. Apparently, $u(0)=0, u(1)=5$. Using knowledge from functional analysis, one can further argue that $u \in C^1[0,1]$ and is the minimum of this problem.

Arzela showed the necessity of equicontinuity and boundedness for compactness while Ascoli established the compactness under equicontinuity and boundedness. Google under Arzela-Ascoli theorem for details.

There are some fundamental results that require completeness. The contraction mapping principle is one and Baire category theorem is another. The latter was first introduced by Baire in his 1899 doctoral thesis. It has wide, and very often amazing applications in all branches of analysis. Some nice applications are available on the web. Google under applications of Baire category theorem for more.

Weierstrass' example is discussed in Hewitt-Stromberg, "Abstract Analysis". An simpler example can be found in Rudin's Principles.

Being unable to locate a single reference containing these three topics, I decide not to name any reference but let you search through the internet.