Banach Contraction mapping principle:

Let (X, d) be a complete metric space. suppose $T : X \to X$ is a contraction map., that is, there is a number c; $0 \leq c < 1$ such that $d(Tx, Ty) \leq cd(x, y)$ for all points x, y. Thus the distance between images is 'smaller' than the distance between original points. In other words T decreases distance. The main point is that distance reduction is by a fixed proportion; not simply that it is reduced.

The theorem says that there is a unique fixed point, that is, a point x^* such that $Tx^* = x^*$. We saw that there can not be two fixed points. To show that there is one, let us start with any point x. Put

$$x_0 = x; \quad x_1 = Tx_0; \quad x_n = Tx_{n-1} \quad n \geq 1.$$

We shall show that (x_n) is a Cauchy sequence.

Then by completeness there is a point x^* such that $x_n \to x^*$. Observe that if $a_n \to a$ then $Ta_n \to Ta$ because

$$d(Ta_n, Ta) \leq cd(a_n, a) \to 0.$$

Thus $Tx_n \to Tx^*$. But $Tx_n = x_{n+1} \to x^*$. Thus $Tx^* = x^*$.

We now show that (x_n) is Cauchy. For $n \geq m$

$$\begin{aligned} d(x_n, x_m) \ &\leq \ cd(x_{n-1}, x_{m-1}) \ \leq \ c^2 d(x_{n-2}, x_{m-2}) \\ &\leq \ \cdots \ \leq \ c^m d(x_0, x_{n-m}) \\ &\leq \ c^m \{d(x_0, x_1) + d(x_1, x_2) + \cdots + d(x_{n-m-1}, x_{n-m})\} \\ &\leq \ c^m d(x_0, x_1)\{1 + c + c^2 + c^3 + \cdots + c^{n-m}\} \\ &\leq c^m d(x_0, x_1) \frac{1}{1-c}. \end{aligned}$$

Let $\epsilon > 0$ be given. Since $c < 1$ choose N so that $c^N d(x_0, x_1)/(1-c) < \epsilon$. Let now $n, m \geq N$. without loss of generality, let us say $n \geq m$. The estimate above shows that $d(x_n, x_m) < \epsilon$. This completes proof.

This is a very useful tool and a powerful technique. We shall discuss some examples.

First let us note that every continuous map of an interval $[a, b]$ to itself has a fixed point; whether it is a contraction or not. This is because if f is the map, then $f(x) - x$ is a continuous function. Since values of f are in $[a, b]$, we see $f(a) - a \geq a - a = 0$ and $f(b) - b \leq b - b = 0$. The intermediate value theorem completes the proof.

However the above argument does not tell that fixed point is unique. In fact it may not be, as for example, the identity map $f(x) = x$ shows. The argument does not tell you how to get a fixed point. If f is a C^1 function (continuously differentiable) then the derivative is bounded. If $|f'| \leq c < 1$ then f is a contraction and so if you start from any point in this interval and keep applying f successively, you will be heading to the fixed point.

Not only that, you get an idea of how close you are to the fixed point. Indeed the estimate obtained above tells you (letting $n \to \infty$) that

$$d(x_m, x^*) \leq c^m \, d(x_0, x_1)/(1 - c).$$

contraction on R:

Consider the function

$$f(x) = \frac{1}{2} \sin(\cos x) + 239$$

Then you see $|f'| \leq 1/2$ and hence it is a contraction on R. Thus there is a unique fixed point. Of course, this is nothing exciting because you see that the values of the function lie in the interval $239 \pm 1/2$. So you can regard f as a map of this interval to itself. But as explained earlier, you have a method to obtain the fixed point.

contractions on R^n:

When is a linear map of R^n a contraction? More generally, let us consider an affine map $Tx = Ax + b$ where $A = (a_{ij})$ is an $n \times n$ matrix and b is a vector in R^n. Obviously whether a map is a contraction or not depends on the metric and here we have several metrics on R^n and under each of them it is complete — all the metrics are equivalent. But let us consider, for illustration, only two of those.

Let us consider the Euclidean distance

$$d(x,y) = \sqrt{\sum (x_i - y_i)^2}.$$

Thus

$$d(Tx,Ty) = \sqrt{\sum_i \left\{ \sum_j a_{ij}(x_j - y_j) \right\}^2} \le \sqrt{\sum_i \left\{ \sum_j a_{ij}^2 \quad \sum_j (x_j - y_j)^2 \right\}}.$$

$$= \sqrt{\sum_{i,j} a_{ij}^2}\, d(x,y).$$

Thus if $\sum\limits_{i,j} a_{ij}^2 < 1$ then the map T is a contraction (in the Euclidean metric) and hence there is a fixed point.

Let us take the distance

$$d(x,y) = \sum_i |x_i - y_i|.$$

Then denoting the column sum $\sum\limits_i |a_{ij}| = c_j$ and $c = \max c_j$ we have

$$d(Tx,Ty) = \sum_i \left| \sum_j a_{ij}(x_j - y_j) \right| \le \sum_i \sum_j |a_{ij}|\ |x_j - y_j|$$

$$= \sum_j c_j |x_j - y_j| \le c d(x,y).$$

Thus if $\max\limits_j \sum\limits_i |a_{ij} < 1$ then T is a contraction in the d_1 metric and hence has a fixed point.

Let us take the distance

$$d(x,y) = \max |x_i - y_i|$$

Then denoting the row sum $r_i = \sum\limits_j |a_{ij}|$ and $r = \max r_i$

$$d(Tx,Ty) = \max_i \left| \sum_j a_{ij}(x_j - y_j) \right| \le \max_i \{ \sum_j |a_{ij}| d(x,y) \} \le r d(x,y)$$

Thus if $\max\limits_i \sum\limits_j |a_{ij}| < 1$ then T is a contraction in the d_∞ metric and hence has a fixed points.

Thus the main point is that if any one of the above conditions holds then there is a unique fixed point. Of course you can consider the d_p metrics too.

inverse function theorem:

Another application of the contraction mapping theorem is for proving the inverse function theorem. We proved it last year. We have an open set $\Omega \subset R^n$ and a C^1 function f defined on Ω to R^n. We are given a point $x_0 \in \Omega$ and are told that the derivative matrix f' at x_0 is non-singular. Then the conclusion is that there is an open set U with $x_0 \in U \subset \Omega$ and an open set $V \subset R^n$ such that f is one-one on U onto V and the inverse map on V onto U is again C^1. Of course there is an expected formula for the derivative of the inverse map.

You can look up that proof. After showing open sets U and V we had a small estimate which is reminiscent of contraction. We used it to conclude that for a point $y \in V$ there can not be two points $x_1, x_2 \in U$ with $f(x_1) = f(x_2) = y$. However, to show that there is a point x at all, we used a hands on calculation involving solution of linear equations. We could have used fixed point theorem. This was done in your calculus III course by Balaji. So we shall not repeat. Actually, you can look up this specific point in our notes, just to get a glimpse of how we grazed (and not used) the fixed point theorem.

differential equations:

We shall now discuss another important application. This application is for solving differential equations. In high school you discussed theory of equations. We have number, do not know what it is, but we know that it satisfies $x^2 - 5x + 6 = 0$. We need to find the number. In this case there is explicit formula, you can 'factorize' this polynomial. Sometimes there was no explicit formula but still you were able to deduce existence of solution using some other rules (Descarte's rule of signs or intermediate value theorem or whatever).

In the study of differential equations, you have a function; you do not know what it is; but you know that the function and its derivatives obey some relation. You need to find the function. Just as in the case of theory of equations it is in general difficult to find explicit formula for the function. Sometimes you can solve explicitly and get a formula for the function. But such a luxury is very rare. We should be satisfied if we know that there is a function at all satisfying the differential equation. We should be more than happy if we know how many such functions are there. If we succeed, we shall try to understand the solution more.

If some one asked you to solve $x'(t) = \sin t$ then it is very easy because the function is $x(t) = -\cos t + c$ for some number c. This was simple because of two reasons. firstly there is no appearance of the function on the right side. Secondly we could integrate the function. You can integrate whenever right side does not involve the function x (or its derivatives etc). Whether you can get a formula or not depends on the right side. For example $x'(t) = \exp\{-t^2\}$ will not allow you to come up with formula. But in these cases you are sure that the integral of the right side is a solution.

Serious problems arise if the right side also involves the function x. For example solve: $x'(t) = \sin x(t)$. We do not even know if there is a solution at all (yes, there is, you will see). There is no need for us to define what is a differential equation and what is meant by a solution. This is reserved for a later course. We consider a specific problem, so you will have no trouble following.

Differential equations arise in several contexts. More important are they in physics. Imagine that in the space in this room at every point there is a force. Of course, when I say force you will ask; on what is it acting. No, it is not yet acting — it has the potential to act. For example if there is a particle at the point P the force at that point will push it in a particular direction. suppose it is pushed to Q, then the force at that point Q will push it to R and the force at the point R will push it to etc and so on.

There is just one subtle point which is very important. I simply said the force at P will act on my particle and push it to Q. Actually when the particle is on its way to Q it will pass through intermediate points, but there are forces at these points and they do their bit as well to push. Thus the particle may actually not reach Q and it may be pushed to Q'. Pause and think about the situation our particle is in. since there are forces at every point, our particle is pushed *at every time instant.* In other words it is continuously pushed around. Life for our particle is not as discrete as I made it out in the earlier para.

So what is the problem. Well, I now place a particle at this point P. Tell me how it travels. Tell me the trajectory or path the particle takes. so you should tell me at every time instant t the position of the particle. We shall not consider in this generality and in three dimensions.

Another situation where these arise is in geometry. Basically, I want to draw a curve and I have been instructed 'how the curve should curve'. More precisely, at every point of the plane, R^2, a vector is given. I was given a starting point P. I should draw my curve starting at the given point P and at any point on the curve, the tangent is as prescribed at the point, remember there are placed vectors at every pint of R^2 and if your curve passes through a point then tangent to your curve at that point should be as suggested.

Now let us make matters precise. We are given an open set $U \subset R^2$ and a point $(t_0, x_0) \in U$. We are given a function $F : U \to R$. We are asked to locate an interval $(t_0 - \delta, t_0 + \delta)$ and a differentiable real valued function x on this interval satisfying two conditions: $x(t_0) = x_0$ (initial condition) and for every t in this interval $x'(t) = F(t, x(t))$ (differential equation).

You see the change of attitude. Our task is 'local'. In a small interval around the given point t_0 we should solve the problem. The idea is that once you do this you can 'continue' from where you 'reached'. Try to make sense of it. It does not concern us now because we are not going to carry it out and find out what is the largest interval on which the solution can be defined. We shall show locally and stop. Observe that $(t, x(t))$ is a point on the curve and we are wanting at that point of the curve the derivative should be same as value of F at that point.

We solve this problem assuming certain conditions. Here is the precise theorem.

Theorem (Picard):
Given
(i) Open set $U \subset R^2$ and a point $(t_0, x_0) \in U$.
(ii) $F : U \to R$ which is continuous and there is a number M such that $|F(t, x) - F(t, y)| \le M|x - y|$ for all points (t, x) and (t, y) in U.
Then *there exists*
an interval $(t_0 - \delta, t_0 + \delta)$ and a real valued differentiable function x on this interval such that its graph lies in U and at every point t in this interval $x'(t) = F(t, x(t))$.

Proof is very simple. fix a closed ball B with centre (t_0, x_0) such that $B \subset U$. Let K be an bound for F on the compact set B. Choose $\delta > 0$ so that

(i) $M\delta < 1$ and

(ii) $[t_0 - \delta, t_0 + \delta] \times [x_0 - K\delta, x_0 + K\delta] \subset B$.

Before we produce the function, let us make an observation that motivates the later considerations.

Solving

$x'(t) = F(t, x(t))$ and $x(t_0) = x_0$ $\quad (\spadesuit)$

is same as solving

$x(t) = x_0 + \int_{t_0}^{t} F(s, x(s))ds.$ $\quad (\clubsuit)$

Indeed, suppose $(\spadesuit)$ holds. Then x being differentiable, must be continuous and F being continuous $F(t, x(t))$ is continuous. In other words x' is continuous. Then rules of integration tell us that

$$x(t) - x(t_0) = \int_{t_0}^{t} x'(t)dt = \int_{t_0}^{t} F(s, x(s))ds$$

which is $(\clubsuit)$.

Conversely suppose $(\clubsuit)$ holds. Then the fundamental theorem of integration gives $(\spadesuit)$.

Consider the space X consisting of functions $x \in C[t_0 - \delta, t_0 + \delta]$ satisfying the following two conditions.

(i) $x(t_0) = x_0$ and

(ii) values of x lie in the interval $[x_0 - K\delta, x_0 + K\delta]$.

In other words the space of continuous functions on the said interval whose value at t_0 is as required and $|x(t) - x_0| \leq K\delta$ for every t in this interval.

The space $C[t_0 - \delta, t_0 + \delta]$ is a complete metric space — exactly like $C[0, 1]$ is complete. Its subset X we are considering is a closed subset. If a sequence of continuous functions satisfying the two conditions converge to a function in the space (so uniform convergence) then the limit function also satisfies the two conditions. Hence X is a complete metric space.

Let us define a map T of X to itself by

$$Tx(t) = x_0 + \int_{t_0}^{t} F(s, x(s)ds.$$

The above analysis suggests that if we find a fixed point of this map then it satisfies our requirements. Thus the proof is completed by showing firstly, that the map T we defined takes X to itself and secondly, it is a contraction.

Since F is continuous and indefinite integral is continuous we see that Tx is continuous function. Actually the indefinite integral is continuous even if integrand is not continuous, but we never proved such a theorem. We however proved that for a continuous function the indefinite integral is continuous. This is all what we need to see Tx is a continuous function.

Also from definition of Tx we see that its value at t_0 is indeed x_0. Further, since x takes values in the interval $[x_0 - K\delta, x_0 + K\delta]$ the graph of x lies in our rectangle $[t_0 - \delta, t_0 + \delta] \times [x_0 - K\delta, x_0 + K\delta]$ and hence the value of x is bounded by K. Thus

$$|Tx(t) - x_0| = \left| \int_{t_0}^{t} F(s, x(s)ds \right| \le K|t - t_0| \le K\delta$$

Thus $Tx \in X$ for each $x \in X$.

Finally take any $x, y \in X$ and $t \in [t_0 - \delta, t_0 + \delta]$. Let us assume $t > t_0$; similar argument applies if $t < t_0$.

$$|Tx(t) - Ty(t)| = \left| \int_{t_0}^{t} \{ \ F(s, x(s)) - F(s, y(s)) \ \} ds \right|$$

$$\le \int_{t_0}^{t} |F(s, x(s) - F(s, y(s)| ds \le \int_{t_0}^{t} M|x(s) - y(s)| ds$$

$$\le \int_{t_0}^{t} M d(x, y) \ ds \le M\delta d(x, y).$$

This shows $d(Tx, Ty) \le (M\delta) d(x, y)$ and since $M\delta < 1$ we have proved that T is indeed a contraction.

This complete the proof.

Recall that a function g of one variable is said to be Lipschitz if there is a number M such that $|g(x) - g(y)| \le M|x - y|$ for all x, y in its domain. The number is called the Lipschitz constant. Thus what we demanded, apart from continuity of F, is that it be Lipschitz in the second variable x for each fixed value for the first variable t. Further the Lipschitz constant does not depend on t, thus the same number M works for all t. This condition is expressed by saying that F is *Lipschitz in the second variable, uniformly in the first variable.*

You might be wondering what has such a theorem got to do with physics, because, the equations you come across there involve second derivatives. Remember, the simplest situation is $F = ma$ or $mx''(t) = F$. Here $x(t)$ is position of particle at time t and $x'(t)$ will give you velocity and $x''(t)$ gives the acceleration. so to know where the particle is you should solve a more complicated equation that we considered. I shall explain the trick.

Suppose you want to solve $x''(t) = -x(t)$. Of course, you know the solutions already. Solving for one function which involves second derivative is transformed to a problem of solving for two functions but involves only first derivative as follows. Solve for two functions x and y with

$$x'(t) = y(t); \qquad y'(t) = -x(t)$$

If you could solve the first problem and get x, take $y = x'$ to see you have a solution for the later problem. Conversely, if you have solution x, y for the second problem then x solves the first problem.

The method for solving for two functions is similar to the above, involves no new ideas but we shall not get into.

integral equations and iterations:

We shall show two other interesting flowers from the fixed point theory garden. These are a little advanced and we need tools like scissors or blade to pluck these flowers; bare hands will not do as in the earlier examples. So we only see and be happy.

We discussed conditions for an affine map to be contraction. More precisely we have an $n \times n$ matrix and an n vector b. We considered the following map on R^n.

$$Tx_i = \sum_j A_{ij} x_j + b_i; \quad i = 1, 2, \cdots, n.$$

Instead of denoting vectors by (x_i) let us denote as $\{x(t) : 1 \leq t \leq n\}$ and accordingly we denote the matrix by $A(s,t)$. Thus the map takes the form

$$Tx(s) = \sum_t A(s,t)x(t) + b(s); \quad 1 \leq s \leq n.$$

Its appearance suggests a natural interpretation. Think of x as a function!

More precisely, consider $C[0,1]$. Suppose you are given a function b in this space. suppose we are given a continuous function A on $[0,1] \times [0,1]$. Consider the problem of finding a function x on $[0,1]$ such that

$$x(s) = \int_0^1 A(s,t)x(t)dt; \quad 0 \leq s \leq 1.$$

Thus if you define Tx as the function on the right side above, then the problem is to find a fixed point of $Tx = x$.

The natural stage for this problem is not the set of continuous functions on $[0, 1]$ but functions x on $[0, 1]$ such that $x^2(t)$ is integrable. This can be solved using Banach fixed point theorem. Such problems are called integral equations, as the appearance itself suggests.

Here is another problem. What is the Cantor set? You are familiar with it. Start with interval $[0, 1]$ and delete middle one-third intervals repeatedly. Here is another way of looking at it.

Consider the two functions on the interval $[0, 1]$ into itself.

$$T_1(x) = \frac{1}{3}x; \qquad T_2(x) = \frac{2}{3} + \frac{1}{3}x.$$

Start with the 'seed' the two points $\{0, 1\}$. If you keep on applying the above maps what do you get?

$$K_0 = \{0;\ 1\}$$

$$K_1 = \{0;\ 1/3;\ 2/3;\ 1\}.$$

$$K_2 = \{0;\ 1/9;\ 2/9;\ 3/9;\ 6/9;\ 7/9;\ 8/9;\ 1\}.$$

What is happening to these sets? They are converging to the Cantor set! You will wonder how that can happen because each of these sets is finite and the limit should, at the best, be countable. Also you will wonder what this has got to do with contraction. Also, this appears like a complicated way of explaining Cantor set that we know so well! Is it of any use at all. These issues are what we explain now.

Consider the space X which consists of all non-empty closed subsets (equivalently, compact subsets) of $[0, 1]$. For example the set $\{0, 1\}$ or any finite set or Cantor set or the interval $[1/3, 1/2]$ etc are all elements of this set X. The set $\{1/n : n = 1, 2, 3, \cdots\}$ is not an element of this set but the set $\{0, 1/n; n = 1, 2, 3, \cdots\}$ is an element of this set.

Hausdorff defined a nice metric on the set X which makes it a complete metric space (actually compact metric). This is a nice metric: Two sets which appear to your eye close are close in this metric. We shall not get into precise definition.

Define the following map T on this space.

$$TK = T_1(K) \cup T_2(K) = \left\{ \frac{1}{3}x; \ \frac{2}{3} + \frac{1}{3}x; \quad x \in K \right\}.$$

It so happens that T is a contraction of this space X. If you start with the set K_0 described above then the successive iterations lead you to K_1, K_2 and so on. It is possible to prove that the sequence $T^n K_0$ converges to C, the Cantor set in the Hausdorff distance. Thus you get Cantor set as a result of iteration of a contraction.

So what is the use. I need to send you a picture of Cantor set. Scanning and sending it takes too much space. I can feed the map T and the seed and instruct iteration. Then the computer can iterate a large number of times and plot the resulting set. This will be an excellent approximation of the Cantor set. Eventhough not all points of the Cantor set are plotted, enough are plotted and your eye believes it is seeing Cantor set!

Of course, you might ask whether the construction process, namely, starting with an interval and deleting middle one-thirds could as well be followed. This process can be iterated a large number of times and the resulting set can be plotted. No matter how many times you iterate you will plot intervals and the picture will not reveal the true nature of Cantor set, after all, Cantor set does not contain any non-trivial interval.

Of course, I illustrated using a trivial set that you are familiar, but most important situation is when you want to describe beautiful designs. There is a beautiful theory behind manufacturing impressive designs.

Compact spaces:

We have been discussing completeness and its consequences. Let us pass to another topic which is equally important.

A metric space (X, d) is compact if the following is true: Given any collection of open sets whose union is X, we can pick finitely many of those open sets whose union equals X. In other words, given a collection of open sets that cover X, there are finitely many of those which also cover X. The first order of business is to relate ti to what we know in R^n.

Theorem 1: X is compact implies every sequence has a limit point.

Suppose a sequence (x_n) has no limit point. Then given any point x there is an open ball B_x such that $x \in B_x$ and which does not contain infinitely many terms of the sequence; that is, there is a stage after which no term of the sequence is in this ball. Consider such a ball for each point. These balls cover X. Pick finitely many of these balls which cover X. But then, there is a stage after which no term of the sequence is in any of these finitely many balls. But those points are in X! This contradiction proves the statement.

Theorem 2: X is compact implies every sequence has a convergent subsequence.

We know that there is a subsequence converging to limit point of the sequence. so this follows form the previous theorem.

Theorem 3: X is compact implies that every Cauchy sequence converges.

If a Cauchy sequence (x_n) has a subsequence that converges to a point x, then the sequence itself converges to x. This is easy and can be seen as follows. Let $n_1 < n_2 < n_3 < \cdots$ be such that

$$x_{n_1}, x_{n_2}, x_{n_3}, \cdots \to x.$$

We show $x_n \to x$. Let $\epsilon > 0$ be given. Fix k so that $d(x_{n_i}, x) < \epsilon/2$ for $i \geq k$. fix $N > n_k$ such that $d(x_n, x_m) < \epsilon/2$ for $m, n \geq N$. Let us take any m larger than N. Pick an n_i larger than N. Then

$$d(x_m, x) \leq d(x_m, x_{n_i}) + d(x_{n_i}, x) \leq \epsilon.$$

To prove the theorem, note that given any Cauchy sequence, the previous theorem says that there is a subsequence which converges. But then as seen above the sequence itself converges.

Theorem 4: X is compact implies the following. For any given $\epsilon > 0$, finitely many ϵ balls cover X. That is, there are finitely many balls whose radius is ϵ and their union equals X.

Here you can interpret balls as open balls or as closed balls, the statement is true.

Given $\epsilon > 0$ take open ball of radius ϵ around each point $x \in X$. Clearly all these open balls cover X. Since open ball is open set, compactness implies finitely many of these balls cover X.

Closed Balls with same centres and same radius ϵ too cover X.

Theorem 5: If X is compact then the following two conditions hold.
(i) X is complete.
(ii) For any given $\epsilon > 0$, finitely many ϵ balls cover X.
Conversely if (i) and (ii) hold then the metric space is compact.

If X is compact, then by Theorem 3 every Cauchy sequence converges and hence the space is complete. Second part is just the previous theorem.

Conversely let us assume that (i) and (ii) hold. We show X is compact. Fix any collection $\mathcal{U}$ of open sets whose union is X. We repeat exactly the same proof that we did in R^n.

Assume that no finite sub collection covers X. Take finitely many closed balls of radius one which cover X. If each one of these can be covered by finitely many sets from $\mathcal{U}$ then surely X can also be covered. Since this is not the case, fix one such ball B_1 which can not be covered by finitely many sets from $\mathcal{U}$.

Take finitely many closed balls of radius $1/2$ which cover X. Take their intersection with the above B_1. If each one of these can be covered by finitely many sets from $\mathcal{U}$ then surely B_1 can also be covered. Since this is not the case, fix one ball B_2 of radius $1/2$ so that $B_1 \cap B_2$ can not be covered by finitely many sets from $\mathcal{U}$.

Thus by induction, we can get a sequence of balls (B_i) such that
For each i, B_i is a closed ball of radius $1/i$ and
for each n, $B_1 \cap B_2 \cap \cdots \cap B_n$ can not be covered by finitely many sets from $\mathcal{U}$.

By completeness and Cantor intersection theorem you get a point $x \in \cap_{i\geq 1} B_i$. Since $\mathcal{U}$ covers X, pick $U \in \mathcal{U}$ with $x \in U$. Pick $\epsilon > 0$ with $B(x,\epsilon) \subset U$. Pick N so that $1/N < \epsilon/4$. Then we claim $B_N \subset U$. To see this first observe that distance between any two points of B_N is at most $2/N$ (go via centre). Also $x \in B_N$. Thus for any $y \in B_N$ we have $d(x,y) \leq 2/N \leq \epsilon/2$. In other words $y \in B(x,\epsilon) \subset U$.

Thus $B_1 \cap B_2 \cap \cdots \cap B_N$ is covered by just one set from $\mathcal{U}$. But our construction says this is not possible. This completes the proof that finitely many sets from $\mathcal{U}$ indeed cover X.

Note that R is complete but not compact. Thus condition (i) alone is not enough to deduce compactness. The metric space $X = (0, 1)$ satisfies condition (ii) but is not compact. Thus condition (ii) alone is not enough to deduce compactness.

Theorem 6: X is compact iff every sequence has a convergent subsequence.

If X is compact then theorem 2 already shows that every sequence has a convergent subsequence.

Conversely, assume that every sequence has a convergent subsequence. We show X is compact. We verify the two conditions of the previous theorem.

To show completeness, take any Cauchy sequence, then hypothesis tells us that there is a convergent subsequence but then the sequence itself converges as observed earlier.

Let $\epsilon > 0$ be given. We need to show that finitely many ϵ balls cover the space. suppose it is false. Take any point x_1. Since $B_1 = B(x_1, \epsilon)$ is not all of X, pick $x_2 \notin B_1$ and let $B_2 = B(x_2, \epsilon)$. since $B_1 \cup B_2$ is not all of X, pick $x_3 \notin B_1 \cup B_2$. proceeding in this way we pick a sequence of points

$$x_1, x_2, x_3, \cdots; \quad x_n \notin \bigcup_{i \leq n-1} B(x_i, \epsilon).$$

Clearly distance between any two points of the sequence is at least ϵ. This sequence has no convergent subsequence. If it has, say converging to p then $B(p, \epsilon/4)$ should contain at least two terms of the sequence, which would mean distance between those two points is at most $\epsilon/2$.

This completes the proof.

We shall now discuss compactness of subsets. Let (X, d) be a metric space. Let $Y \subset X$. Recall that we can regard Y itself as a metric space; metric being the restriction of d to $Y \times Y$.

Theorem 7: The following two statements are equivalent.

(i) Given any family of subsets of Y which are open in Y there are finitely many whose union equals Y.

(ii) Given any family of open subsets of X whose union includes Y, there are finitely many of those whose union includes Y.

The importance of this theorem is the following. When you re talking about compact subsets of a pace X, any of the above two statements can be used as definition. The statement (i) forgets the background and says that the metric space (Y, d) is compact. On the other hand statement (ii) does not forget the background and states in terms of the open sets of X, does not refer to open subsets of Y at all.

To prove the theorem we only need to observe the following;

(*) A set $U \subset Y$ is open in Y iff there is a set $V \subset X$ which is open in X such that $U = V \cap Y$.

Assume, for a moment, truth of the above statement. We prove the theorem as follows. Let (i) hold. Let (V_α) be a family of sets open in X whose union includes Y. Then $U_\alpha = V_\alpha \cap Y$ are open in Y and cover Y and finitely many U_α cover Y and then the corresponding V_α cover Y.

Conversely let (ii) hold and (U_α) be a collection of sets open in Y whose union is Y. For each α pick a set V_α which is open in X and $U_\alpha = V_\alpha \cap Y$. these (V_α) cover Y and so finitely many of them cover Y and the corresponding U_α cover Y.

Proof of (*) is routine. Suppose V is open in X and let $U = V \cap Y$. Need to show U is open in Y. Let $y \in U$. So $y \in V$. Since V is open in X there is a ball in X, say $B_X(y, r) \subset V$. The ball in Y of radius r, that is $B_Y(y, r)$ is nothing but $B_X(y, r) \cap Y$ and is hence contained in U. This shows U is open.

Conversely, let U be an open set in Y. We need to exhibit a set V open in X such that $U = V \cap Y$. For each $y \in U$ we get a ball in Y, say, $B_Y(y, r) \subset U$. This r depends on the point y, we did not show it in the notation. Let $B_X(y, r)$, the ball in X with the same centre be denoted by A_y. Thus remember $A_y \cap Y = B_Y(y, r)$. Union of all these balls A^y be denoted by V. Then V is open in X and $V \cap Y = U$.

This completes the proof.

You should carefully understand the above proof. It is actually trivial, but you can say so only if you understood it.

continuity:

So far we have been taking about sets — closed, open, compact, connected, first category, second category etc.

We shall now study functions between metric spaces.

Let (X, d) and (Y, ρ) be two metric spaces and $f : X \to Y$. We say that f is continuous if it preserves convergence. That is, $x_n \to x$ in X implies that $f(x_n) \to f(x)$ in Y.

Remember our main idea in taking up metric spaces is the feeling that if we have a concept of 'how close things are' we can do at least a part of calculus that deals with convergence and continuous functions.

After all $x_n \to a$ meant that x_n is getting closer and closer to a. Further, our idea of continuous function on R to R is that: when x is close to a then $f(x)$ should be close to $f(a)$. This was made precise by saying that whenever $x_n \to a$ then $f(x_n) \to f(a)$.

We shall now explore continuous functions and their properties.