Notes for 30 Mar (Thursday)

1 Recap

1. Gave several examples of equicontinuous functions. One way to ensure equicontinuity is to have a family of continuously differentiable functions whose derivatives $|f'(x)| \leq M$ where M is independent of x and f.

2. For countable subsets of arbitrary metric spaces we proved that pointwise bounded sequences of functions have convergent subsequences.

3. We proved that if K is compact, then continuous f_n converge to f uniformly implies that f_n are equicontinuous.

4. If K is a compact subset of any metric space (equivalently, K is a compact metric space), f_n are continuous, pointwise bounded and equicontinuous, then f_n are uniformly bounded and have a uniformly convergent subsequence. (The Arzela-Ascoli theorem.)

2 The Weierstrass approximation theorem

This beautiful theorem is interesting for two reasons - The statement of the theorem is itself pleasant, but more importantly there are two proofs of this theorem, each of which illustrates a very important idea.

Suppose you want to store a continuous function on $[0,1]$ on a computer. You need to give a finite description of it. If it is $\sin(x)$, you can simply write the formula (mathematica for example, can manipulate it). If it is piecewise constant, simply store the finitely many values (approximately by rationals). Otherwise you are in trouble. One hope is to perhaps approximate the function by storing its values (approximately) on 1000 equally spaced points. But this is also not good enough.

The Weierstrass approximation theorem approximates every continuous function by polynomials. Indeed,

Theorem 1. *If $f : [a,b] \to \mathbb{C}$ is a continuous function, then there exist polynomials $P_n(x)$ converging uniformly to $f(x)$. If f is real, the P_n may be taken to be real.*

Without loss of generality we may assume that $[a,b]$ is $[0,1]$. As mentioned before, there are two proofs. One of them relies on probability. The other is in Rudin and is an analytic proof. Let's do the probabilistic proof first. It is due to Bernstein.

Proof. Suppose we play the following "game". You are given a biased coin whose probability of falling on its head is x (and on its tail is $1-x$). Suppose you throw it n times and it falls on its head k times, you are paid $f(k/n)$ rupees where f is the given function. (If it is complex, the money you get paid is in your head, just like the money you receive from the government in a railway accident.) So the question is : On an average, if you throw the coin lots of times, how much money will you win ?

Of course, on an average, the coin will fall roughly x fraction of the times tossed. Therefore you should expect around $f(x)$ rupees. On the other hand, the change of the coin falling k times on its head is $\binom{n}{k}x^k(1-x)^{n-k}$. Therefore the amount you will win exactly is $\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}f(k/n)$. These two should be roughly equal when n is large. Thus f is approximated by a polynomial. Moreover, if f is real, of course the polynomial has real coefficients. This can actually be made rigorous using the (weak) law of large numbers but we will give an elementary proof.

Let us be more precise :

$$|\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}f(k/n)-f(x)| = |\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}f(k/n)-\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}f(x)|$$
$$= |\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}(f(k/n)-f(x))|$$

Since f is continuous on a compact set, it is uniformly continuous and hence $|f(x)-f(y)|<\epsilon/2$ when $|x-y|<\delta$. Thus we can choose n to be so large that $\frac{1}{n}<\delta$. Then $|f(k/n)-f(x)|<\epsilon/2$ for $|k/n-x|<\delta$. Therefore,

$$|\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}(f(k/n)-f(x))| < \frac{\epsilon}{2}+2M\sum_{|k/n-x|\geq\delta}\binom{n}{k}x^k(1-x)^{n-k} \tag{1}$$

for sufficiently large n. For such k, the binomial coefficients are small. Indeed, let us calculate how much k/n differs from x on an average :

$$\sum_{k=0}^{n}\left(\frac{k}{n}-x\right)^2\binom{n}{k}x^k(1-x)^{n-k} = x^2-2x\sum_{k=0}^{n}\frac{k}{n}\binom{n}{k}x^k(1-x)^{n-k}+\sum_{k=0}^{n}\frac{k^2}{n^2}\binom{n}{k}x^k(1-x)^{n-k}$$
$$= x^2-2x\sum_{k=1}^{n}\binom{n-1}{k-1}x^k(1-x)^{n-k}+\frac{1}{n^2}\sum_{k=0}^{n}(k(k-1)+k)\binom{n}{k}x^k(1-x)^{n-k}$$
$$= x^2-2x^2+\frac{1}{n^2}(n(n-1)x^2+nx) = \frac{x(1-x)}{n}$$

The left hand side of the above equation is greater $\sum_{|k/n-x|\geq\delta}\binom{n}{k}x^k(1-x)^{n-k}\delta^2$. Therefore, we see that $\sum_{|k/n-x|\geq\delta}\binom{n}{k}x^k(1-x)^{n-k} < \frac{x(1-x)}{n\delta^2}$. This goes to 0 as $n\to\infty$. Hence $|P_n(x)-f(x)|<\epsilon$ for large n. □

The second proof illustrates an idea in analysis borrowed from physics.

Proof. Using $g(x) = f(x) - f(0) - x(f(1) - f(0))$ we may assume without loss of generality that $f(0) = f(1) = 0$. We now extend f to be 0 outside $[0, 1]$. Thus f is uniformly continuous on the entire real line.

The rough idea is that $\int_{-1}^{1} \delta(t - x) f(t) dt = f(x)$ for the Dirac delta "function". If you approximate the Dirac delta by polynomials, you will be in great shape.

There is a systematic way to approximate the Dirac delta. First, take $(1 - x^2)$. This is 0 when $x = \pm 1$ and reaches its maximum (just like the Dirac delta) at $x = 0$. Now if you take $(1 - x^2)^n$, then it is very small away from $x = 0$ when n is large. Unfortunately, it is just 1 at $x = 0$ (unlike the Dirac delta). So let $Q_n(x) = c_n(1 - x^2)^n$ where c_n is chosen so that $\int_{-1}^{1} Q_n(x) dx = 1$, i.e., $c_n = \dfrac{1}{\int_{-1}^{1} (1 - x^2)^n dx}$.

Now let $P_n(x) = \int_{0}^{1} f(t) Q_n(t - x) dt = \int_{-1}^{1} f(x + t) Q_n(t) dt$. We hope that $|P_n(x) - f(x)| < \epsilon$ for all $x \in [0, 1]$ when n is large.

$$
\begin{aligned}
|P_n(x) - f(x)| &= |\int_{-1}^{1} Q_n(t) f(t + x) - \int_{-1}^{1} Q_n(t) dt f(x)| \\
&= |\int_{-1}^{1} Q_n(t)(f(x + t) - f(x)) dt|
\end{aligned}
$$

Suppose n is large. When t is very small, $Q_n(t)$ is huge, but the integral over such small t is still small. If t is not so small, then $Q_n(t)$ is extremely small and f is bounded. So the right hand side is expected to be small when n is large. Let us analyse $Q_n(t)$ when t is small and when t is large. For this, we need to know how c_n behaves. (*To be cont'd...*) □