Chapter 4

Limits and Continuity

4.1 Mathematical Modelling

Few people study or pursue mathematics out of interest in mathematics itself. Most do so because mathematics is needed for their main interest.

Probably the most common application of mathematics is *modelling*, phenomena, processes or complex systems in the natural sciences, in engineering,in economics and elsewhere.

There are two principal requirements for a mathematical model:

- *Accuracy*;
- *Computability*.

These are usually antagonistic, confronting us with the optimisation problem of maximising (ease of) computation while minimising loss of accuracy. Our approach will be to start with the simplest possible functions to compute, increasing the complexity only to the extent needed to improve accuracy to the desired level.

In this course, we restrict ourselves to models based on real-valued functions of a real variable.

We adopt the view that the simplest functions to compute are the *polynomial functions*,

$$p : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \sum_{j \in \mathbb{N}} a_j x^j := a_0 + a_1 x + a_2 x^2 + \cdots + a_j x^j + \cdots$$

with only finitely many $a_j \neq 0$, as these can be computed using a finite number of elementary arithmetic operations.

For a non-zero polynomial p, we takes its *degree*

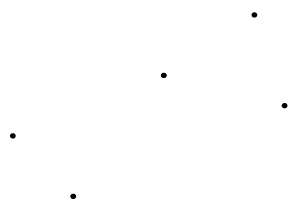
$$\deg(p) := \min\{k \in \mathbb{N} \mid a_j = 0 \text{ for } j > k\}$$

as a measure of computational complexity.

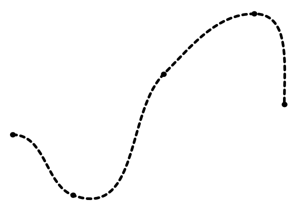
Thus, the simplest polynomials are those of degree 0, the *constant* polynomials.

Our approach will be to find the polynomial of a given degree, which best fits the available data — if there is such a polynomial — and only seeking greater accuracy at the cost of increased complexity if the polynomial function is no sufficiently accurate. We shall see that to achieve each increase in accuracy, we shall need to make more stringent assumptions on the functions we approximate, and so on the processes we can model in this manner.

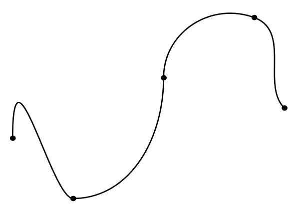
We suppose we are give a set of experimental data, in expressed graphically in the form of points in the plane. The working hypothesis is that these points lie on the graph of a function, and our task is to find that function, or, equivalently, the curve of "best fit" passing through the data points.



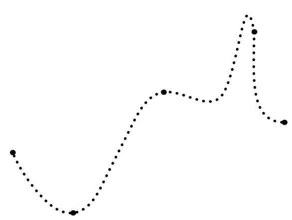
Possible curves include



and



and



As there is no *a priori* way of solving this problem, we approach it by first treating each data point separately, to find the curve of best fit locally, and then seek to splice these local curves into a global one.

The working hypothesis we use is that the curve in question is the graph of a real valued function of a real variable, $f\colon \mathbb{R} \to \mathbb{R}$.

We seek to approximate it using polynomial functions

$$p_n(x) = c_0 + c_1 x + \cdots + c_n x^n$$

We regard the *degree*, n, of the polynomial function p_n as a measure of its complexity and seek to find the polynomial approximation of lowest degree which is most accurate.

We shall see that this places increasingly strict constraints on the function f with increasingly stringent demand for accuracy.

We assume we have a point $(a, f(a))$ and try to find our simplest polynomial approximate f near $x = a$, namely

$$p_0(x) = c_0$$

In order for this to fit at a the curve which is the graph of f, we must have $p_0(a) = f(a)$.

Thus $p_0(x) := f(a)$ is the only possible polynomial function of degree 0 to fit our curve at $x = a$.

The question arises

> When is this p_0 a good approximation to f near $x = a$?

We first make the question more precise and formulate it mathematically.

While our approximation is accurate at $x = a$, it is of little use if it is only accurate there. We would like to know that as long as x does not stray too far from a, $p_0(x)$ will still be close to $f(x)$.

To facilitate a mathematical form of this, we reverse the rôles of f and p_0, since $p_0(x)$ is close $f(x)$ if and only if $f(x)$ is close to $p_0(x)$, and while we know precisely all the values of p_0, we do not know those of f.

Informally, we say that the function f approximates the number ℓ near a, if and only if

> **Given any tolerance about ℓ, there is a deviation about a such that whenever x lies within the deviation about a, $f(x)$ is within the tolerance of ℓ.**

In such a case, we say that $f(x)$ *tends to* ℓ *as* x *tends to* a.

Before we formulate this mathematically, we observe that the actual value of f is irrelevant. Indeed, we do not even need f to be defined at a.

Example 113. Choose $a \in \mathbb{R}$. The function

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \begin{cases} 100 & \text{for } x = a \\ 0 & \text{otherwise} \end{cases}$$

Plainly, f is a good approximation to 0 near a, but it not at a.

To formulate our informal notion precisely, we note that we have a measure of distance between two real numbers, u, v, namely $|u - v|$, the absolute value of the difference between them.

Denoting the tolerance about ℓ by ε and the deviation from a by δ we arrive at a rigorous definition.

Definition 114. Let X be a set of real numbers. Then the function $f : X \to \mathbb{R}$ *tends to* ℓ *as* x *tends to* a, or, equivalently, f has *limit* ℓ *as* x *tends to* a if and only if

> **Given $\varepsilon > 0$, there is a $\delta > 0$ with $|f(x) - \ell| < \varepsilon$, if $0 < |x - a| < \delta$.**

We write

$$f(x) \to \ell \;\text{ as }\; x \to a$$

or

$$\lim_{x \to a} f(x) = \ell.$$

While this definition is straightforward, its direct application can be anything easy. For given a function $f\colon X \to \mathbb{R}$ it is first necessary to find a plausible candidate for ℓ.

When $a \in X$, an immediate candidate is $f(a)$. Example 113 shows that this does not always provide the answer. Those functions for which $f(a)$ does provide the answer, are precisely the *continuous functions*.

Definition 115. Let X be a set of real numbers. The function $f\colon \mathbb{R} \to \mathbb{R}$ is *continuous at* $a \in X$ if and only if

$$\lim_{x \to a} f(x) = f(a),$$

or, equivalently,

Given any $\varepsilon > 0$, there is a $\delta > 0$ with $|f(x) - f(a)| < \varepsilon$, whenever $|x - a| < \delta$.

f is said to be *continuous* if and only if it is continuous at every $a \in X$.

Another suggestive formulation of continuity at a is that

$$f\left(\lim_{x \to a} x\right) = \lim_{x \to a} f(x).$$

We return to our question as to when polynomial function $p_0(x) = f(a)$ is a good approximation to $f(x)$ near a.

The error arising from using p_0 instead of f at x is $|f(x) - f(a)|$, and we want the to be as accurate as possible the closer we get to a. This is achieved whenever $|f(x) - f(a)| \to 0 \quad$ a$x \to a$, or, equivalently,

$$\lim_{x \to a} |f(x) - f(a)| = 0$$

Thus we see that our polynomial function $p_0(x) = f(a)$ is a good approximation to the function $f\colon X \to \mathbb{R}$ if and only if f is continuous to a.

Since limits are central to our considerations, we turn to studying them.

4.2 Properties of Limits

We begin by extending Definition 114 to incorporate either the independent variable or the value of the function is unbounded.

Definition 116. Take $X \subseteq \mathbb{R}$, a function $f\colon X \to \mathbb{R}$ and $a \in X$. Then

(i) $f(x) \to \infty$ (resp. $-\infty$) as $x \to a$ if and only if given any $K \in \mathbb{R}$, there is a $\delta > 0$ such that $f(x) > K$ (resp.$f(x) < K$) whenever $0 < |x| < \delta$.

(ii) $f(x) \to \ell$ as $x \to \infty$ (resp. $-\infty$) if and only if given any $\varepsilon > 0$, there is an $L \in \mathbb{R}$ such that $|f(x) - \ell| < \varepsilon$ for all $x \in X$ with $x > L$ (resp. $x < L$).

(iii) $f(x) \to \infty$ (resp. $-\infty$ as $x \to \infty$ (resp. $-\infty$) if and only if given any $K \in \mathbb{R}$, there is an $L \in \mathbb{R}$ such that $f(x) > K$ (resp.$f(x) < K$) for all $x \in X$ with $x > L$ (resp. $x < L$).

It is not uncommon to abuse notation by writing

$$\lim_{x\to a} f(x) = \pm\infty \quad \text{or} \quad \lim_{x\to\pm\infty} f(x) = \pm\infty$$

in cases (i) and (iii) above. We shall only do so when the circumlocution needed to avoid the abuse obstructs the clarity of a statement.

Theorem 117. *Take functions $f, g, h\colon \mathbb{R} \to \mathbb{R}$ and suppose that*

$$f(x) \to \ell,\ g(x) \to m \text{ as } x \to a \quad \text{and} \quad h(y) \to n \text{ as } y \to \ell$$

where we allow $\pm\infty$ for a, ℓ, m, n. Then

$$\begin{aligned}
(f+g)(x) &\to \ell + m && \text{as } x \to a && \text{unless } \ell = \infty \text{ and } m = -\infty \text{ or vice versa;}\\
(fg)(x) &\to \ell m && \text{as } x \to a && \text{unless } \ell = 0 \text{ and } m = \pm\infty \text{ or vice versa;}\\
(h \circ f)(x) &\to n && \text{as } x \to a && \text{if } h \text{ is continuous at } \ell \text{ whenever } \ell \in \mathbb{R};\\
\frac{1}{f(x)} &\to \frac{1}{\ell} && \text{as } x \to a && \text{whenever } \ell \neq 0\\
\ell &\leq m && && \text{whenever } f(x) \leq g(x) \text{ for all } x \text{ “near” } a
\end{aligned}$$

Proof. We first prove the results for the case $a, \ell, m, n \in \mathbb{R}$.

Take $\varepsilon > 0$.

Then $\varepsilon' := \frac{\varepsilon}{2} > 0$, and so, since $f(x) \to \ell$ and $g(x) \to m$ as $x \to a$, there are $\delta_1, \delta_2 > 0$ with $|f(x) - \ell| < \varepsilon'$, whenever $0 < |x - a| < \delta_1$ and $|g(x) - m| < \varepsilon'$, whenever $0 < |x - a| < \delta_2$.

Put $\delta := \min\{\delta_1, \delta_2\}$ and suppose $0 < |x - a| < \delta$. Then

$$\begin{aligned}
|(f+g)(x) - (\ell + m)| &= |f(x) + g(x) - \ell - m|\\
&\leq |f(x) - \ell| + |g(x) - m|\\
&< \varepsilon' + \varepsilon'\\
&= \varepsilon,
\end{aligned}$$

showing that $(f + g)(x) \to \ell + m$ as $x \to a$.

Take $\zeta > 0$ and choose δ such that $|f(x) - \ell|, |g(x) - m| < \zeta$ whenever $0 < |x - a| < \delta$.

Since $\big||g(x)| - |m|\big| \leq |g(x) - m|$, if $0 < |x - a| < \delta$, then

$$|g(x)| < |m| + \varepsilon,$$

and so

$$\begin{aligned}
|(fg)(x) - \ell\, m| &= |f(x)g(x) - \ell\, m|\\
&\leq |f(x) - \ell|\,|g(x)| + |\ell|\,|g(x) - g(a)|\\
&\leq \zeta(m + \zeta) + \ell\zeta\\
&= \zeta^2 + (\ell + m)\zeta
\end{aligned}$$

By the quadratic formula, $\zeta^2 + (\ell + m)\zeta \leq \varepsilon$ if and only if

$$0 \leq \zeta \leq \frac{-(\ell+m) + \sqrt{(\ell+m)^2 + 4\varepsilon}}{2}$$

So, given $\varepsilon > 0$, chose a positive $\delta < \dfrac{-(\ell+m) + \sqrt{(\ell+m)^2 + 4\varepsilon}}{2}$.

By the above, if $0 < |x - a| < \delta$, then $|(fg)(x) - \ell\, m| < \varepsilon$. So $(fg)(x) \to \ell\, m$ as $x \to a$.

Since h is continuous at ℓ, there is a $\zeta > 0$ with $|h(y) - n| < \varepsilon$ whenever $|y - \ell| < \zeta$.

Because $f(x) \to \ell$ as $x \to a$, there is a $\delta > 0$ with $|f(x) - \ell| < \zeta$ whenever $0 < |x - a| < \delta$.

So, given x with $0 < |x - a| < \delta$, $|f(x) - \ell| < \zeta$, whence $|(h \circ f)(x) - n| = |\big(f(x)\big) - n| < \varepsilon$, showing that $(h \circ f)(x) \to n$ as $x \to a$.

Suppose $\ell \neq 0$.

Take $\varepsilon > 0$.

Then $\zeta := \min\{\dfrac{\varepsilon|\ell|^2}{2}, \frac{|\ell|}{2}\} > 0$.

Choose $\delta > 0$ such that for all x with $0 < |x - a| < \delta$, $|f(x) - \ell| < \zeta$. Then

$$\big||f(x)| - |\ell|\big| \leq |f(x) - f(a)| < \zeta,$$

whence

$$0 < \frac{|\ell|}{2} < |f(x)| < \frac{3|\ell|}{2}$$

so that

$$0 < \frac{2}{3|\ell|} < \frac{1}{|f(x)|} < \frac{2}{|\ell|}.$$

Hence, if $0 < |x - a| < \delta$,

$$\begin{aligned} \left|\frac{1}{f(x)} - \frac{1}{\ell}\right| &= \frac{|f(x) - \ell|}{|f(x)|\,|\ell|} \\ &< \frac{2\zeta}{|\ell|^2} \\ &\leq \varepsilon, \end{aligned}$$

showing that $\frac{1}{f(x)} \to \frac{1}{\ell}$ as $x \to a$.

Suppose that $\ell > m$. Put $\frac{m-\ell}{2}$.

Then there is a $\delta > 0$ such that for all x with $0 < |x - a| < \delta$,

$$\ell - \varepsilon < f(x) < \ell + \varepsilon = m - \varepsilon < g(x) < m + \varepsilon,$$

showing that $f(x) > g(x)$ for $0 < |x - a| < \delta$.

□

Observation 118. A convenient mnemonic for Theorem 117 is

$$\begin{aligned} \lim(f+g) &= (\lim f) + (\lim g) \\ \lim(fg) &= (\lim f)(\lim g) \\ \lim(h \circ f) &= \lim h \\ \lim\left(\frac{1}{f}\right) &= \frac{1}{\lim f} \end{aligned}$$

and limits preserve inequality. However this mnemonic should be used mindfully of the additional conditions in Theorem 117.

Corollary 119. *Take functions $f, g, h\colon \mathbb{R} \to \mathbb{R}$ such that f, g are continuous at a and h is continuous at $f(a)$. Then $f+g, fg$ and $h \circ f$ are all continuous at a. $\frac{1}{f}$ is continuous at a if $f(a) \neq 0$.*

4.3 Properties of Continuous Functions

We have seen that in order to be able to approximate a function $f\colon \mathbb{R} \to \mathbb{R}$ near a, by means of a polynomial function of degree 0, f must be continuous at a.

We turn to investigating properties of continuous functions.

Our first observation is that a function can be continuous at exactly on point.

Example 120. The function

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \begin{cases} x & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

is continuous at 0 and only at 0.

To see this, take $a \neq 0$ and put $\epsilon := \frac{|a|}{2}$. For each $\delta > 0$, there is a rational number, q, and an irrational number, s, with $a < q, s < a + \delta$.

If $a \in \mathbb{Q}$, then $|f(a) - f(s)| = |a| > \varepsilon$.

If $a \notin \mathbb{Q}$, then $|f(a) - f(q)| = |q| > |a| > \varepsilon$.

Hence, $f$is not continuous at a.

On the other hand, given $\varepsilon > 0$, put $\delta := \varepsilon$. Take x with $|x - 0| < \delta$. Then

$$|f(x) - f(0)| = |f(x)| \leq |x| < \delta = \varepsilon,$$

showing that f is continuous at 0.

Lemma 121. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous at a. Then there is an $r > 0$ such that $\{f(x) \mid a - r < x < a + r\}$ is a bounded set of real numbers.*

Proof. Since f is continuous at a there is an $r > 0$ with $f(a) - 1 < f(x) < f(a) + 1$, whenever $a - r < x < a + r$. □

Theorem 122 (Intermediate Value Theorem). *Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous. Take $a, b \in \mathbb{R}$ with $a < b$. For each y between $f(a)$ and $f(b)$ there is an $x \in [a, b]$ with $f(x) = y$.*

Proof. If $f(a) = f(b)$, there is nothing to prove.

We consider the case $f(a) < f(b)$ and leave the case $f(a) > f(b)$ to the reader as an exercise.

Suppose that there is a $y \in \,]f(a), f(b)[\setminus \operatorname{im}(f)$.

Put $A := \{x \in [a, b] \mid f(t) < y \text{ for all } t \in [a, x]\}$.

By definition, b is an upper bound for A.

Since $a \in A$, $A \neq \emptyset$.

Hence A has a supremum, say s. Plainly $s \in [a, b]$.

Then either $f(s) < y$ or $f(s) > y$.

Put $\varepsilon := |f(s) - y|$, and take $\delta > 0$.

If $f(s) < y$, then $\varepsilon = y - f(s)$.

Since $s = \sup A$, it follows from the definition of A that there is an $x \in [s, +\delta[$ with $f(x) > y$. Then

$$\begin{aligned}
|f(x) - f(s)| &= f(x) - f(s) && \text{as } f(x) > y > f(s) \\
&= f(x) - y + y - f(s) \\
&= f(x) - y + \varepsilon \\
&> \varepsilon && \text{as } f(x) > y,
\end{aligned}$$

whence f is not continuous at s.

If $f(s) > y$, then $\varepsilon = f(s) - y$.

Choose $x \in A$ with $s - \frac{\delta}{2} < x$.

By the definition of A, $f(x) < y$. Then

$$\begin{aligned}
|f(x) - f(s)| &= f(s) - f(x) && \text{as } f(x) < y < f(s) \\
&= f(s) - y + y - f(x) \\
&= \varepsilon + y - f(x) \\
&> \varepsilon && \text{as } f(x) < y,
\end{aligned}$$

whence f is not continuous at s. □

Theorem 123. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous and take $a, b \in \mathbb{R}$ with $a < b$. Then $\{f(x) \mid a \leq x \leq b\}$ is bounded.*

Proof. Put $A := \{x \in [a, b] \mid f \text{ is bounded on } [a, x]\}$.

Since $a \in A$, $A \neq \emptyset$. Plainly b is an upper bound for A. Thus A has a supremum, $s \in [a, b]$.

But f is continuous on $[a, b]$. So, by Theorem 123, there is an $r > 0$ with f bounded on $]s - r, s + r[\cap [a, b]$, and hence on $[a, s + r]$.

Since $S = \sup A$, this is only possible if $]s, s + r[\cap [a, b] = \emptyset$,. Thus $s = b$. □

Theorem 124 (Extreme Value Theorem). *Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous. If $a < b$, then $\{f(x) \mid a \leq x \leq b\}$ has both a minimum and a maximum.*

Proof. We show that $f([a, b]) := \{f(x) \mid a \leq x \leq b\}$ has a maximum, and leave as exercise for the reader the proof that $f([a, b])$ has a minimum.

By Theorem 123, $f([a, b])$ is bounded. It is non-empty because $f(a) \in f([a, b])$. Hence it has a supremum, s.

Suppose that $s \notin f([a, b])$. Then $s - f(x) > 0$ for all $x \in [a, b]$, hence, by Corollary 119,

$$g\colon [a, b] \longrightarrow \mathbb{R}, \quad x \longmapsto \frac{1}{1 - f(x)}$$

is a continuous function.

By Theorem 123, $g([a, b])$ is bounded. Hence, there is a $K > 0$ with $g(x) < K$ for all $x \in [a, b]$. As $g(x) > 0$ for all $x \in [a, b]$

$$\frac{1}{K} < \frac{1}{g(x)} = s - f(x),$$

or, equivalently,

$$f(x) < s - \frac{1}{K} < s,$$

for all $x \in [a, b]$, which contradicts the fact that $s = \sup\{f(x) \mid a \leq x \leq b\}$. □

Theorem 125. *Let $f\colon [a, b] \to \mathbb{R}$ be continuous and injective. Then f is monotonic.*

Proof. Since $f\colon [a, b] \to \mathbb{R}$ is injective, $f(a) \neq f(b)$.

We consider the case $f(a) < f(b)$ and leave the case $f(a) > f(b)$ to the reader as an exercise.

Take $x \in]a, b[$. Since f is injective, $f(x) \neq f(a), f(b)$.

Suppose $f(x) < f(a)$. Then $f(x) < f(a) < f(b)$.

Since f is continuous on $[x, b]$, there is, by Theorem 122, a $u \in [x, b]$ with $f(u) = f(a)$.

Since $a < x$, this contradicts the injectiveness of f, whence $f(x) > f(a)$.

Now take y with $a < x < y < b$. We repeat the argument above to show that $f(x) < f(y)$.

Since f is injective, $f(y) \neq f(x), f(b)$.

Suppose $f(y) < f(x)$. Then $f(y) < f(x) < f(b)$.

Since f is continuous on $[x, b]$, there is, by Theorem 122, a $u \in [x, b]$ with $f(u) = f(x)$.

Since $a < x$, this contradicts the injectiveness of f, whence $f(y) > f(x)$.

If $f(y) > f(b)$, then $f(a) < f(b) < f(y)$, so that, by the Intermediate Value Theorem, there is an $x \in [a, y]$ with $f(x) = f(b)$, contradicting the injectiveness of f.

Thus, f is (strictly) monotonic increasing on $[a, b]$. □