MATH 209: PROOF OF EXISTENCE / UNIQUENESS THEOREM FOR FIRST ORDER DIFFERENTIAL EQUATIONS

INSTRUCTOR: STEVEN MILLER

ABSTRACT. We highlight the proof of Theorem 2.8.1, the existence / uniqueness theorem for first order differential equations. In particular, we review the needed concepts of analysis, and comment on what advanced material from Math 301 / 305 (real analysis) is needed. We include appendices on the Mean Value Theorem, the Intermediate Value Theorem, and Mathematical Induction. The only result we need which is non-elementary and is not proved in these notes is the Lebesgue Dominated Convergence Theorem. This is a major result, and allows us to interchange a limit and an integral; however, it should be possible to prove the special case we need elementarily (the proof is left as an exercise for the reader). There are numerous problems throughout the handout so that you can test your understanding of the material if desired.

1. STATEMENT

Theorem 1.1. *Let f and $\partial f/\partial y$ be continuous functions on the rectangle $R = [-a, a] \times [-b, b]$. Then there is an $h \leq a$ such that there is a unique solution to the differential equation $dy/dt = f(t, y)$ with initial condition $y(0) = 0$ for all $t \in (-h, h)$.*

Following the textbook, we have elected to simplify notation and not state the theorem in the greatest generality. We have performed two translations so that we assume the time interval is centered at 0 and the y values are centered at 0. There is no loss in such generality. To see this, consider instead the equation $du/d\tau = g_1(\tau, u(\tau))$ with $u(\tau_0) = u_0$. Clearly this is the most general such first order equation; we now show that we may transform this into the form of Theorem 1.1. Let $v(\tau) = u(\tau) - u(\tau_0)$. Note $v(\tau_0) = u(\tau_0) - u(\tau_0) = 0$, and as $dv/d\tau = du/d\tau$ we see $dv/d\tau = g_1(\tau, v(\tau) + u(\tau_0)) = g_2(\tau, v(\tau))$. This shows that there is no loss in generality in assuming the initial value is zero. A similar argument shows we may change the time variable to assume the initial time is zero.

Date: February 26, 2009.

Exercise 1.2. *Find the time change of variables to prove that we may assume the time variable is centered at 0.*

2. Analysis pre-requisites

We need several results from Real Analysis, which we now collect below.

Lemma 2.1. *Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function on a finite interval $[\alpha, \beta]$. Then there is some M such that $|g(x)| \leq M$. If instead $g : \mathbb{R}^2 \to \mathbb{R}^2$ is a continuous function on a finite rectangle $[\alpha, \beta] \times [\gamma, \delta]$, then there is an M such that $|g(x, y)| \leq M$.*

A nice feature of many analysis proofs is that the exact value of M doesn't matter, instead what is important is that there is some finite value for M which works. As an example, consider the function $g(x) = e^{-x^2} - x^4 + x^2 \cos(2x)$ on the interval $[0, 2]$. We have

$$|g(x)| \ \leq \ |e^{-x^2}| + |x^4| + |x^2| \cdot |\cos(2x)|; \tag{1}$$

this is a *very* wasteful way to find an upper bound for g, but it will yield one. The largest the first term can be is 1, the largest the second is $2^4 = 16$, and the largest the last is $2^2 = 4$; thus $|g(x)| \leq 1+16+4 = 21$.

Exercise 2.2. *The trivial estimate above isn't off by much; the actual maximum value is about 19. Determine the optimal value. Unfortunately if you try to use calculus and find the critical points, you end up with an extremely difficult problem to solve. You'll have to use Newton's method or divide and conquer. Alternatively, if you can show the first derivative is positive for $x > 1$ you know the maximum value is at the endpoint. One must be careful as we care about the maximum of the absolute value, and thus you have to break the analysis into cases where g is positive and negative. This is one reason why we often just estimate crudely.*

Definition 2.3 (Absolutely and Conditionally Convergent Series)**.** *We say a series is absolutely convergent if $\sum_{n=0}^{\infty} |a_n|$ converges to some finite number a; if $\sum_{n=0}^{\infty} |a_n|$ diverges but $\sum_{n=0}^{\infty} a_n$ converges to a finite number a, we say the series is conditionally convergent.*

Example 2.4. *The series $\sum_{n=0}^{\infty} 1/2^n$ is absolutely convergent, while the series $\sum_{n=0}^{\infty} (-1)^n/n$ is only conditionally convergent.*

Lemma 2.5 (Comparison Test). *Let $\{b_n\}_{n=0}^{\infty}$ be a sequence of non-negative numbers whose sum converges; this means $\sum_{n=0}^{\infty} b_n = b < \infty$ (and this immediately implies $\lim_{n\to\infty} b_n = 0$). If $\{a_n\}_{n=0}^{\infty}$ is another sequence of real numbers such that $|a_n| \le b_n$ then $\sum_{n=0}^{\infty} a_n$ converges to some finite number a.*

Exercise 2.6. *If $\sum_{n=0}^{\infty} a_n$ converges absolutely, show $\lim_{n\to\infty} a_n = 0$.*

Remark 2.7. *In Lemma 2.5, we don't need $|a_n| \le b_n$ for all n; it suffices that there is some N such that for all $n \ge N$ we have $|a_n| \le b_n$. This is because the convergence or divergence of a series only depends on the tail of the sequence; we can remove finitely many values without changing the limiting behavior (convergence or divergence).*

Example 2.8. *We know $\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$ if $|r| < 1$. Thus the series $a_n = (-1)^n/n!$ converges as $|a_n| \le (1/2)^n$ for $n \ge 1$.*

Exercise 2.9. *Prove the geometric series formula: if $|r| < 1$ then $\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$.*

Other useful series to know are the p-series. Let $C > 0$ be any real number and let $p > 0$. Then $\sum_{n=1}^{\infty} \frac{C}{n^p}$ converges if $p > 1$ and diverges if $p \le 1$.

Exercise 2.10. *Prove $\sum_{n=1}^{\infty} \frac{1}{n^2+2n+5}$ converges.*

Exercise 2.11. *Prove $\sum_{n=1}^{\infty} x^n/n!$ converges for all x (or at least for $|x| < 1$).*

Theorem 2.12 (Lebesgue's Dominated Convergence Theorem). *Let f_n be a sequence of continuous functions such that (1) $\lim_{n\to\infty} f_n(x) = f(x)$ for some continuous function f, and (2) there is a non-negative continuous function g such that $|f_n(x)|$ and $|f(x)|$ are at most $g(x)$ for all x and $\int_{-\infty}^{\infty} g(x)dx$ is finite. Then*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)dx \;=\; \int_{-\infty}^{\infty} \lim_{n\to\infty} f_n(x)dx \;=\; \int_{-\infty}^{\infty} f(x)dx. \qquad (2)$$

We have stated this result with significantly stronger conditions than is necessary, as these are the conditions that hold in our problem of interest.

Exercise 2.13. *Let $f_n(x) = 0$ if $x \le n$, $n(x-n)$ if $n \le x \le n+1$, $n(n+2-x)$ if $n+1 \le x \le n+2$, and 0 otherwise; thus $f_n(x)$ is a triangle of height n and width 2 centered at $n+1$. Show that for any x, $\lim_{n\to\infty} f_n(x) = 0$. Why can't we use Lebesgue's Dominated Convergence Theorem to conclude that $\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)dx = 0$?*

Exercise 2.14. *Prove Theorem 2.12.*

The last result we need is the Mean Value Theorem; we give a proof in Appendix A (the proof uses the Intermediate Value Theorem, which we also prove).

Theorem 2.15 (Mean Value Theorem (MVT))**.** *Let $h(x)$ be differentiable on $[a,b]$, with continuous derivative. Then*

$$h(b)-h(a) \;=\; h'(c)\cdot(b-a), \qquad c\in[a,b]. \tag{3}$$

Remark 2.16 (Application of the MVT)**.** *For us, one of the most important applications of the Mean Value Theorem is to bound functions (or more exactly, the difference between a function evaluated at two nearby points). For example, let us assume that f is a continuously differentiable function on $[0,1]$. This means that the derivative f' is continuous, so by Lemma 2.1 there is an M so that $|f'(w)|\le M$ for all $w\in[0,1]$. Thus we can conclude that $|f(x)-f(y)|\le M|x-y|$. This is because the Mean Value Theorem gives us the existence of a $c\in[0,1]$ such that $f(x)-f(y)=f'(c)(x-y)$; taking absolute values and noting $|f'(c)|\le \max_{0\le w\le 1}|f'(w)|$, which by Lemma 2.1 is at most M, yields the claim.*

Exercise 2.17. *Let $f(x) = e^{x^2-4} - x^2\sin(x^2+2x) + \frac{x+1}{x^2+5}$. Prove $|f(x)-f(y)|\le 6|x-y|$ whenever $x,y\in[0,2]$.*

3. Step 1 of the Proof of Theorem 1.1

In the proof of Theorem 1.1 (see the textbook), we use Picard's iteration method to construct a sequence of functions $\phi_n(t)$ by setting $\phi_0(t)=0$ and

$$\phi_{n+1}(t) \;=\; \int_0^t f\left(s,\phi_n(s)\right)ds. \tag{4}$$

Note that $\phi_n(0)=0$ for all n, which is good (as we are trying to solve the differential equation $dy/dt=f(t,y)$ with initial condition $y(0)=0$).

We want to prove two facts: first, that $\phi_n(t)$ exists for all n, and second that it is continuous. If $\phi_n(t)$ exists for some n, then $\phi_{n+1}(t)$ exists as well. This is because

$$\phi_{n+1}(t) \;=\; \int_0^t f(s,\phi_n(s))ds, \tag{5}$$

and the integral of a continuous function is continuous (regard $f(s, \phi_n(s))$ as some new function, say $g(s)$, and now we can use our results from first year calculus). The only problem is that we want $\phi_{n+1}(t)$ to always lie in the interval $[-b, b]$.

Recall that we are trying to solve the differential equation $dy/dt = f(t, y)$ for $t \in [-a, a]$ and $y \in [-b, b]$. As f is continuous, by Lemma 2.1 there is an M such that $|f(t, y)| \leq M$ for all $t \in [-a, a]$ and all $y \in [-b, b]$. If we restrict to $t \in [-h, h]$ for $h \leq \min(b/M, a)$, then the integral

$$\int_0^t f(s, \phi_n(s))ds \tag{6}$$

is at most $M|t| \leq Mh \leq b$. We see now why we restricting to $t \in [-h, h]$ is potentially needed; this ensures that $\phi_{n+1}(t)$ takes on values in $[-b, b]$.

4. Step 2 in the Proof of Theorem 1.1

We want to show that $\lim_{n\to\infty} \phi_n(t)$ exists for all t (*and is continuous!*). We write $\phi_n(t)$ as

$$\phi_n(t) \ = \ \sum_{k=1}^{n} \left(\phi_n(t) - \phi_{n-1}(t)\right), \tag{7}$$

remembering that $\phi_0(t) = 0$. Thus

$$\phi_3(t) \ = \ (\phi_1(t) - 0) + (\phi_2(t) - \phi_1(t)) + (\phi_3(t) - \phi_2(t))\,; \tag{8}$$

you might recall that this is a telescoping sum. These sums are often easy to analyze (and play a role in some of the proofs of the Fundamental Theorem of Calculus).

We now show that $\lim_{n\to\infty} \phi_n(t)$ exists for all t. *We can do this for any $r < 1$ by making sure h is sufficiently small (remember we have restricted to studying only $t \in (-h, h)$).* Let us fix some $t \in (-h, h)$. Assume we could show that there is some $r < 1$ such that $|\phi_k(t) - \phi_{k-1}(t)| \leq r^k$ for all k. Then for this t the limit exists by the Comparison Test (Lemma 2.5). (To use the comparison test, we let $b_n = r^n$ and $a_n = \phi_n(t) - \phi_{n-1}(t)$, and note that $\lim_{n\to\infty} \phi_n(t) = \sum_{n=1}^{\infty} a_n$.

Thus we are reduced to showing that there is an $r < 1$ with

$$|\phi_k(t) - \phi_{k-1}(t)| \ \leq \ r^k. \tag{9}$$

This will follow from using the Mean Value Theorem to estimate the integrals for ϕ_k and ϕ_{k-1} and Mathematical Induction (for a review of Mathematical Induction, see Appendix B. Recalling that $\phi_k(t) = \int_0^t f(s, \phi_{k-1}(s))ds$ and similarly for $\phi_{k-1}(s)$, we find

$$\phi_k(t) - \phi_{k-1}(t) \ = \ \int_0^t \left[f(s, \phi_{k-1}(s)) - f(s, \phi_{k-2}(s))\right] ds. \qquad (10)$$

We now apply the Mean Value Theorem (Theorem 2.15) to the function $g(y) = f(s, y)$. We take our two points to be $\phi_{k-1}(s)$ and $\phi_{k-2}(s)$. Thus there is some point $c_k(s)$ between $\phi_{k-1}(s)$ and $\phi_{k-2}(s)$ such that

$$g\left(\phi_{k-1}(s)\right) - g\left(\phi_{k-2}(s)\right) \ = \ g'(c_k(s)) \cdot \left(\phi_{k-1}(s) - \phi_{k-2}(s)\right); \qquad (11)$$

we chose to write the point as $c_k(s)$ to remind ourselves that it depends on k and s (in particular, we are doing this for *every* s in the integral). Noting that $g'(y) = \partial f/\partial y$, we see we have shown

$$\phi_k(t) - \phi_{k-1}(t) \ = \ \int_0^t \frac{\partial f}{\partial y}\left(s, c_k(s)\right) \cdot \left(\phi_{k-1}(s) - \phi_{k-2}(s)\right) ds. \qquad (12)$$

We now proceed by induction. We assume that we have already shown $|\phi_{k-1}(s) - \phi_{k-2}(s)| \le r^{k-1}$ for $s \le t$. We substitute this into (12), and use Lemma 2.1 to bound $\partial f/\partial y$ by M and find

$$|\phi_k(t) - \phi_{k-1}(t)| \ \le \ \int_0^t Mr^{k-1} ds \ = \ Mtr^{k-1} \ \le \ Mhr^{k-1}. \qquad (13)$$

As long as we choose h so that $Mh < r$ (i.e., $h < r/M$), then we obtain the desired result!

Exercise 4.1. *Give the details for the proof by induction. In particular, do the basis case and the inductive step carefully.*

We now want to show that $\phi(t) = \lim_{n\to\infty} \phi_n(t)$ is continuous. (It better be continuous, as we want it to be the solution to the differential equation, and if it isn't continuous then it can't be differentiable!). To show ϕ is continuous, we must show that given any $\epsilon > 0$ there is a $\delta > 0$ such that $|t_2 - t_1| < \delta$ implies $|\phi(t_2) - \phi(t_1)| < \epsilon$. For notational convenience assume $t_1 < t_2$. We have

$$\begin{aligned} \phi(t_2) - \phi(t_1) \quad &= \quad \lim_{n\to\infty} \phi_n(t_2) - \lim_{n\to\infty} \phi_n(t_1) \\ &= \quad \lim_{n\to\infty} \left(\phi_n(t_2) - \phi_n(t_1)\right) \\ &= \quad \lim_{n\to\infty} \int_{t_1}^{t_2} f(s, \phi_n(s)) ds \qquad (14) \end{aligned}$$

(the last line follows from the fact that $\phi(t_1)$ is an integral from 0 to t_1 while $\phi_n(t_2)$ is an integral from 0 to t_2. By Lemma 2.1, there is an M

such that $|f(s,y)| \leq M$. Thus

$$|\phi(t_2) - \phi(t_1)| \ \leq \ \int_{t_1}^{t_2} M ds \ = \ M|t_2 - t_1| \ \leq \ M\delta; \qquad (15)$$

therefore as long as we choose $\delta < \epsilon/M$ we see that $|\phi(t_2) - \phi(t_1)| < \epsilon$.

5. Step 3 in the Proof of Theorem 1.1

In this step we show that the limit function $\phi(t) = \lim_{n\to\infty} \phi_n(t)$ satisfies the differential equation $dy/dt = f(t,y)$ with initial condition $y(0) = 0$ (ie, taking $y(t) = \phi(t)$ gives a solution). From construction, it is clear that $\phi(0) = 0$ as each $\phi_n(0) = 0$. The difficulty is showing that $d\phi/dt = f(s,\phi)$. To see this, we argue as follows:

$$\begin{aligned} \phi(t) &= \lim_{n\to\infty} \phi_n(t) \\ &= \lim_{n\to\infty} \int_0^t f(s, \phi_{n-1}(s)) ds. \end{aligned} \qquad (16)$$

We want to move the limit inside the integral; this can be done because the conditions of Lebesgue's Dominated Convergence Theorem (Theorem 2.12) are met (all functions are continuous, and we may take $g(x) = Mh$ to be the bounding function required by the theorem). Thus

$$\begin{aligned} \phi(t) &= \int_0^t \lim_{n\to\infty} f(s, \phi_{n-1}(s)) ds \\ &= \int_0^t f(s, \lim_{n\to\infty} \phi_{n-1}(s)) ds, \end{aligned} \qquad (17)$$

where the last step (moving the limit inside the function) follows from the fact that f is continuous in each variable. Thus we have shown

$$\phi(t) \ = \ \int_0^t f(s, \phi(s)) ds, \qquad (18)$$

and all functions are continuous. Therefore the Fundamental Theorem of Calculus now yields $\phi'(t) = f(s, \phi(t))$.

6. Step 4 in the Proof of Theorem 1.1

The last result to be shown is that the solution is unique. The proof of this is similar to Step 2. We assume there is another solution $\psi(t)$ and we find

$$\phi(t) - \psi(t) \ = \ \int_0^t \left(f(s,\phi(t)) - f(s,\psi(t))\right) ds. \tag{19}$$

If the two functions are not the same, then there is an $\epsilon > 0$ such that, for some t, $|\phi(t) - \psi(t)| > \epsilon$.

Let

$$m \ = \ \max_{0 \le t \le h} |\phi(x) - \psi(x)| \tag{20}$$

and let M be a bound for $\partial f/\partial y$. Using the Mean Value Theorem we find

$$|\phi(t) - \psi(t)| \ \le \ \int_0^t M\, |\phi(s) - \psi(s)|\, ds \ \le \ M|t|m \ \le \ Mhm. \tag{21}$$

If we choose $h < \epsilon/2mM$, this implies that for all $t < h$, $|\phi(t) - \psi(t)| < \epsilon/2$, which contradicts the fact that there was some t where the difference was at least ϵ.

Appendix A. Proof of the Mean Value Theorem

We will use the Intermediate Value Theorem to prove the Mean Value Theorem.

Theorem A.1 (Intermediate Value Theorem (IVT))**.** *Let f be a continuous function on $[a,b]$. For all C between $f(a)$ and $f(b)$, there exists a $c \in [a,b]$ such that $f(c) = C$. In other words, all intermediate values of a continuous function are obtained.*

Proof. Here is a sketch of a proof using the method of Divide and Conquer. Without loss of generality, assume $f(a) < C < f(b)$. Let x_1 be the midpoint of $[a,b]$. If $f(x_1) = C$ we are done. If $f(x_1) < C$, we look at the interval $[x_1, b]$. If $f(x_1) > C$ we look at the interval $[a, x_1]$.

In either case, we have a new interval, call it $[a_1, b_1]$, such that $f(a_1) < C < f(b_1)$, and the interval has size half that of $[a,b]$. Continuing in this manner, constantly taking the midpoint and looking at the appropriate half-interval, we see one of two things may happen.

First, we may be lucky and one of the midpoints may satisfy $f(x_n) = C$. In this case, we have found the desired point c.

Second, no midpoint works. Thus, we divide infinitely often, getting a sequence of points x_n. This is where rigorous mathematical analysis is required.

We claim the sequence of points x_n converge to some number $X \in (a,b)$. We have an infinite sequence of intervals $\{I_n\}_{n=1}^{\infty}$ such that $I_{n+1} \subset I_n$ and the length of I_{n+1} is half that of I_n. Clearly there cannot be two points in the intersection of all the I_n's. (If there were two points, say α and β, then all points between α and β would also be in the intersection of all these intervals; this is impossible as I_n just isn't big enough once n is so large that the length of I_n is less than $|\beta - \alpha|$.) Thus we are reduced to showing there is one point in the intersection of all these intervals, or equivalently that the x_n's converge to some point X. This is where some rigorous mathematical analysis is required.

Clearly the limit X can't be an endpoint. We keep getting smaller and smaller intervals (of half the size of the previous and contained in the previous) where $f(x) < C$ at the left endpoint, and $f(x) > C$ at the right endpoint. By continuity at the point X, eventually $f(x)$ must be close to $f(X)$ for x close to X.

If $f(X) < C$, then eventually the right endpoint cannot be greater than C; if $f(X) > C$, eventually the left endpoint cannot be less than C. Thus, $f(X) = C$. □

Exercise A.2. *Prove that the sequence of points x_n converges to an $X \in (a,b)$.*

Proof of the Mean Value Theorem. To prove the Mean Value Theorem, it suffices to prove a special case known as Rolle's Theorem, namely that if f is differentiable on $[a,b]$ and $f(a) = f(b) = 0$, then there exists a $c \in [a,b]$ such that $f'(c) = 0$.

To see why it suffices to show Rolle's theorem is true, consider the function

$$h(x) \;=\; f(x) - \frac{f(b)-f(a)}{b-a}(x-a) - f(a). \tag{22}$$

Note $h(a) = f(a) - f(a) = 0$ and $h(b) = f(b) - (f(b) - f(a)) - f(a) = 0$. Thus, the conditions of Rolle's Theorem are satisfied for $h(x)$, and there is some $c \in [a,b]$ such that $h'(c) = 0$. But

$$h'(c) \quad = \quad f'(c) - \frac{f(b)-f(a)}{b-a}. \tag{23}$$

Rewriting yields $f(b) - f(a) = f'(c) \cdot (b-a)$. Thus, it is sufficient to prove Rolle's Theorem to prove the MVT.

We now prove Rolle's theorem. Without loss of generality, assume $f'(a)$ and $f'(b)$ are non-zero (if either were zero, we would be done).

Multiplying $f(x)$ by -1 if needed, we may assume $f'(a) > 0$.

Case 1: $f'(b) < 0$: As $f'(a) > 0$ and $f'(b) < 0$, the Intermediate Value Theorem, applied to $f'(x)$, asserts that all intermediate values are attained. As $f'(b) < 0 < f'(a)$, this implies the existence of a $c \in (a, b)$ such that $f'(c) = 0$.

Case 2: $f'(b) > 0$: $f(a) = f(b) = 0$, and the function f is increasing at a and b. If x is real close to a, then $f(x) > 0$ because $f'(a) > 0$. This follows from the fact that

$$f'(0) \quad = \quad \lim_{x \to 0} \frac{f(x) - f(0)}{x}. \qquad (24)$$

As $f'(0) > 0$, the limit is positive. As the denominator is positive for $x > 0$, the numerator must be positive. Thus, $f(x)$ must be greater than $f(0)$ for small x.

Similarly, $f'(b) > 0$ implies $f(x) < f(b) = 0$ for x near b.

Therefore, the function $f(x)$ is positive for x slightly greater than a and negative for x slightly less than b. If the first derivative were always positive, then $f(x)$ could never be negative as it starts at 0 at a. This can be seen by again using the limit definition of the first derivative to show that if $f'(x) > 0$, then the function is increasing near x. See the next section for more details.

Thus, the first derivative cannot always be positive. Either there must be some point $y \in (a, b)$ such that $f'(y) = 0$ (and we are then done!) or $f'(y) < 0$. By the Intermediate Value Theorem, as 0 is between $f'(a)$ (which is positive) and $f'(y)$ (which is negative), there is some $c \in (a, y) \subset [a, b]$ such that $f'(c) = 0$. □

Appendix B. Proofs by Induction

Assume for each positive integer n we have a statement $P(n)$ which we desire to show is true. $P(n)$ is true for all positive integers n if the following two statements hold:

- **Basis Step:** $P(1)$ is true;
- **Inductive Step**: whenever $P(n)$ is true, $P(n + 1)$ is true.

This technique is called **Proof by Induction**, and is a very useful method for proving results. The reason the method works follows from basic logic. We assume the following two sentences are true:

$$\begin{aligned} P(1) &\text{ is true} \\ \forall n \geq 1, P(n) \text{ is true implies } P(n+1) &\text{ is true.} \end{aligned} \quad (25)$$

Set $n = 1$ in the second statement. As $P(1)$ is true, and $P(1)$ implies $P(2)$, $P(2)$ must be true. Now set $n = 2$ in the second statement. As $P(2)$ is true, and $P(2)$ implies $P(3)$, $P(3)$ must be true. And so on, completing the proof. Verifying the first statement the **basis step** and the second the **inductive step**. In verifying the inductive step, note we assume $P(n)$ is true; this is called the **inductive assumption**. Sometimes instead of starting at $n = 1$ we start at $n = 0$, although in general we could start at any n_0 and then prove for all $n \geq n_0$, $P(n)$ is true.

We give three of the more standard examples of proofs by induction, and one false example; the first example is the most typical.

B.1. **Sums of Integers.** Let $P(n)$ be the statement

$$\sum_{k=1}^{n} k \;=\; \frac{n(n+1)}{2}. \quad (26)$$

Basis Step: $P(1)$ is true, as both sides equal 1.
Inductive Step: Assuming $P(n)$ is true, we must show $P(n+1)$ is true. By the inductive assumption, $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$. Thus

$$\begin{aligned} \sum_{k=1}^{n+1} k &= (n+1) + \sum_{k=1}^{n} k \\ &= (n+1) + \frac{n(n+1)}{2} \\ &= \frac{(n+1)(n+1+1)}{2}. \end{aligned} \quad (27)$$

Thus, given $P(n)$ is true, then $P(n+1)$ is true.

Exercise B.1. *Prove*

$$\sum_{k=1}^{n} k^2 \;=\; \frac{n(n+1)(2n+1)}{6}. \quad (28)$$

Find a similar formula for the sum of k^3.

Exercise B.2. *Show the sum of the first n odd numbers is n^2, i.e.,*

$$\sum_{k=1}^{n} (2k-1) \;=\; n^2. \quad (29)$$

Remark B.3. *We define the empty sum to be 0, and the empty product to be 1. For example,* $\sum_{n\in\mathbb{N},n<0} 1 = 0$.

B.2. **Divisibility.** Let $P(n)$ be the statement 133 divides $11^{n+1} + 12^{2n-1}$.

Basis Step: A straightforward calculation shows $P(1)$ is true: $11^{1+1} + 12^{2-1} = 121 + 12 = 133$.
Inductive Step: Assume $P(n)$ is true, i.e., 133 divides $11^{n+1} + 12^{2n-1}$. We must show $P(n+1)$ is true, or that 133 divides $11^{(n+1)+1} + 12^{2(n+1)-1}$. But

$$\begin{aligned} 11^{(n+1)+1} + 12^{2(n+1)-1} &= 11^{n+1+1} + 12^{2n-1+2} \\ &= 11 \cdot 11^{n+1} + 12^2 \cdot 12^{2n-1} \\ &= 11 \cdot 11^{n+1} + (133 + 11)12^{2n-1} \\ &= 11\left(11^{n+1} + 12^{2n-1}\right) + 133 \cdot 12^{2n-1}. \end{aligned} \tag{30}$$

By the inductive assumption 133 divides $11^{n+1} + 12^{2n-1}$; therefore, 133 divides $11^{(n+1)+1} + 12^{2(n+1)-1}$, completing the proof.

Exercise B.4. *Prove* 4 *divides* $1 + 3^{2n+1}$.

B.3. **The Binomial Theorem.** We prove the Binomial Theorem. First, recall that

Definition B.5 (Binomial Coefficients)**.** *Let n and k be integers with $0 \le k \le n$. We set*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{31}$$

Note that $0! = 1$ and $\binom{n}{k}$ is the number of ways to choose k objects from n (with order not counting).

Lemma B.6. *We have*

$$\binom{n}{k} = \binom{n}{n-k}, \qquad \binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k} \tag{32}$$

Exercise B.7. *Prove Lemma B.6.*

Theorem B.8 (The Binomial Theorem)**.** *For all positive integers n we have*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k. \tag{33}$$

Proof. We proceed by induction.
Basis Step: For $n = 1$ we have

$$\sum_{k=0}^{1} \binom{1}{k} x^{1-k} y^k \;=\; \binom{1}{0} x + \binom{1}{1} y \;=\; (x+y)^1. \tag{34}$$

Inductive Step: Suppose

$$(x+y)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k. \tag{35}$$

Then using Lemma B.6 we find that

$$\begin{aligned}
(x+y)^{n+1} &= (x+y)(x+y)^n \\
&= (x+y) \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \\
&= \sum_{k=0}^{n} \binom{n}{k} x^{n+1-k} y^k + \binom{n}{k} x^{n-k} y^{k+1} \\
&= x^{n+1} + \sum_{k=1}^{n} \left\{ \binom{n}{k} + \binom{n}{k-1} \right\} x^{n+1-k} y^k + y^{n+1} \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} x^{n+1-k} y^k.
\end{aligned} \tag{36}$$

This establishes the induction step, and hence the theorem. □

B.4. **False Proofs by Induction.** Consider the following: let $P(n)$ be the statement that in any group of n people, everyone has the same name. We give a (false!) proof by induction that $P(n)$ is true for all n!

Basis Step: Clearly, in any group with just 1 person, every person in the group has the same name.

Inductive Step: Assume $P(n)$ is true, namely, in any group of n people, everyone has the same name. We now prove $P(n+1)$. Consider a group of $n+1$ people:

$$\{1, 2, 3, \ldots, n-1, n, n+1\}. \tag{37}$$

The first n people form a group of n people; by the inductive assumption, they all have the same name. So, the name of 1 is the same as the name of 2 is the same as the name of 3 ... is the same as the name of n.

Similarly, the last n people form a group of n people; by the inductive assumption they all have the same name. So, the name of 2 is the same as the name of 3 ... is the same as the name of n is the same as the name of $n+1$. Combining yields everyone has the same name! Where is the error?

If $n = 4$, we would have the set $\{1, 2, 3, 4, 5\}$, and the two sets of 4 people would be $\{1, 2, 3, 4\}$ and $\{2, 3, 4, 5\}$. We see that persons 2, 3 and 4 are in both sets, providing the necessary link.

What about smaller n? What if $n = 1$? Then our set would be $\{1, 2\}$, and the two sets of 1 person would be $\{1\}$ and $\{2\}$; there is no overlap! The error was that we assumed n was "large" in our proof of $P(n) \Rightarrow P(n+1)$.

Exercise B.9. *Show the above proof that $P(n)$ implies $P(n+1)$* is *correct for $n \geq 2$, but fails for $n = 1$.*

Exercise B.10. *Similar to the above, give a false proof that any sum of k integer squares is an integer square, i.e., $x_1^2 + \cdots + x_n^2 = x^2$. In particular, this would prove all positive integers are squares as $m = 1^2 + \cdots + 1^2$.*

Remark B.11. *There is no such thing as* Proof By Example. *While it is often useful to check a special case and build intuition on how to tackle the general case, checking a few examples is not a proof. For example, because $\frac{16}{64} = \frac{1}{4}$ and $\frac{19}{95} = \frac{1}{5}$, one might think that in dividing two digit numbers if two numbers on a diagonal are the same one just cancels them. If that were true, then $\frac{12}{24}$ should be $\frac{1}{4}$. Of course this is* not *how one divides two digit numbers!*