Introductory Analysis

Jacob Denson

November 23, 2016

Table Of Contents

Chapter 1

Differential Equations

Here we describe some general techniques for solving differential equations. The goal is simple – we are given an arbitrary vector field $v : \mathbf{R}^n \to \mathbf{R}^n$, and we must find the family of all curves c for which $v \circ c = c'$ holds on the domain of the curve. More generally, we are given a function $F : \mathbf{R}^{n+1} \to \mathbf{R}^m$ (which can be seen as a system of m equations), and we wish to find real-valued differentiable functions f such that

$$F(f(t), \ldots, f^{(n)}(t), t) = 0$$

Geometrically, we must find a curve $c : \mathbf{R} \to \mathbf{R}^{n+1}$ such that

$$c_1'(t) = c_2 \quad c_2'(t) = c_3(t) \quad \ldots \quad c_{n-1}'(t) = c_{n-1}(t)$$

$$F(c_1(t), c_2(t), \ldots, c_{n-1}(t), c_n'(t), c_{n+1}(t)) = 0$$

$$c_{n+1}'(t) = 1$$

Thus we think of c_{n+1} as the time variable, and the other variables as derivatives of another function. Normally we try and find such a solution such that $c_1(0) = x_0, c_2(0) = x_0', \ldots, c_n(0) = x_0^{(n)}, c_{n+1}(0) = t_0$. If F is C^1 and everywhere non-singular, we can use the inverse function theorem to locally write

$$c_{n+1}'(t) = f(c_1(t), \ldots, c_n(t))$$

around $(x_0, \ldots, x_0^{(n)}, t_0)$, which gives us a local vector field around $(c_1, \ldots, c_n)$ defining the same set of solutions as F, so if we assume F is non-singular

we have no real loss of generality in proofs of the local properties of solutions in assuming a vector field defines the differential equation. We will also use Leibnitz notation in certain cases, for instance writing

$$f(x,y)dx + g(x,y)dy = 0$$

For the simple differential equation corresponding to

$$F(t,x,x') = f(t,x) + g(t,x)x'$$

which is non-singular provided that for no (t,x,x') we have

$$\partial_t f(t,x,x') = \partial_t g(t,x)x' = \partial_x f(t,x) = \partial_x g(t,x)x' = g(t,x) = 0$$

simultanously hold.

The most basic type of differential equation is of the form $f'(x) = g(x)$, where g is continuous. The solution family is then found by applying the fundamental theorem of calculus, writing

$$f(x) = f(0) + \int_0^x g(t)$$

and if we let $f(0)$ over all real numbers, we obtain all solutions to the equation. In general, if we find a single function with $f'(x) = g(x)$, then the set of all solutions to the equation is of the form $f(x) + A$.

More advanced techniques are required for more general formulas. A differential equation is separable if it can be rearranged to the form

$$h(y)dy + f(x)dx = 0$$

Any solution to this equation is also a solution to the equation

$$h(f(x))f'(x) = g(x)$$

Applying the chain rule, in reverse, we find

$$h(f(0)) + \int_{f(0)}^{f(x)} h(t) = \int_0^x g(t)$$

Which hopefully simplifies the equation to a more simpler form. If we write the equation in Leibnitz notation

$$\frac{dy}{dx} = g(x)/f(y)$$

3

Then the strategy of separability can be obtained by (formally) writing

$$f(y)dy = g(x)dx$$

And then integrating

$$\int f(y)dy = \int g(x)dx$$

Example. *Consider the equation*

$$e^{f(x)}f'(x) = x(x-2)$$

Integrating, we find

$$e^{f(x)} - e^{f(0)} = \int_{f(0)}^{f(x)} e^t = \int_0^x t(t-2) = x^3/3 - x^2$$

Hence

$$f(x) = \log(e^{f(0)} + x^3/3 - x^2)$$

An equation $f : \mathbf{R}^n \to \mathbf{R}$ is **homogenous of degree** n if $f(tv) = t^n f(v)$. A differential equation

$$M(x,y)dy + N(x,y)dx = 0$$

is homogenous if M and N are homogenous of the same order n. If we let $y = xv$, then $dy = vdx + xdv$, and so

$$xM(x,xv)dv + [vM(x,xv) + N(x,y)]dx = 0$$

Now

$$xM(x,xv) = x^{n+1}M(1,v)$$

$$vM(x,xv) + N(x,xv) = x^n[vM(1,v) + N(1,v)]$$

which separates the variables of N, so we can write

$$\frac{M(1,v)}{vM(1,v) + N(1,v)}dv + \frac{dx}{x}$$

which is a separable equation.

Example.
$$(x^2 + y^2)dx - 2xydy = 0$$

Both coefficients are homogenous of degree 2, so if we let $y = xv$, we find

$$x^2[1 - v^2]dx - 2x^3vdv$$

Which can be rewritten

$$2\frac{vdv}{1 - v^2} = \frac{dx}{x}$$

Hence

$$\log\left(\frac{1 - v_0^2}{1 - v^2}\right) = \log(x/x_0)$$

$$v = \pm\sqrt{1 - \frac{x_0(1 - v_0^2)}{x}}$$

$$y = \pm x\sqrt{1 - \frac{1 - y_0^2}{xx_0}}$$

The differential of a function $f : \mathbf{R}^2 \to \mathbf{R}$ is

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$$

Any differential $adx + bdy$ can be written in this form, as long as the domain is simply connected and $\partial_x b = \partial_y a$. Then if $f(x, c(x)) = A$ for all a, we find

$$a + bc' = \partial_x f + \partial_y f c'(x) = 0$$

So c satisfies the differential equation.

Example. *The differential equation*

$$(1 - 2xy)dx + (4y^3 - x^2)dy = 0$$

can be expressed $df = 0$, where

$$f(x, y) = x - x^2 y + y^4$$

So the solutions the equation satisfy the algebraic equation

$$x - x^2 c(x) + c^4(x) = A$$

5

If a differential equation

$$a\,dx + b\,dy$$

is not exact, we may be able to find f such that $f\,a\,dx + f\,b\,dy$ is exact. f is known as an integrating factor for the equation.

1.1 Linear Equations

Suppose that $v(x_0) = 0$. Then, locally around x_0, we may linearize, writing

$$v_i(x - x_0) = \sum_{j=1}^{n} \frac{\partial v_i}{\partial x_j} x_j$$

The resulting equation is the best linear approximation to v. For instance, linearize the pendulum equation $v(x) = -\sin x$ around a point $n\pi$ gives us the vector field $w(x - n\pi) = (-1)^{n+1} x$
$dy/dx = -\sin x \; dy = -\sin x\,dx \; y - y_0 = \cos(x)-$

Chapter 2

The Exponential Function

We now provide a brief interlude on the subject on the powers of a number. For any complex number x, we may consider its powers x^n, for $n > 0$. Two incredibly useful formulas hold

$$x^{n+m} = x^n x^m \quad (x^n)^m = x^{nm}$$

which can be used to reduce arbitrary calculations of powers to questions about the addition and multiplication of integers. In generalizing the power function to more general numeric systems, we would like these same formulas hold. First, we would like to define x^0. Since we would then have $x^n = x^{n+0} = x^n x^0$, we are forced to define $x^0 = 1$, for all x. Since we would like to have

$$x^{-n} x^n = x^{n-n} = x^0 = 1$$

we are then forced to define x^{-n} to be the unique multiplicative inverse of x^n. These notions hold in any field, and do not require any analysis to define. A more careful analysis is required for generalization to rational numbers. Since

$$(x^n)^q = x^{nq} = x^p$$

we must define x^n to be a q'th root of x^p. The problem here is that q'th roots are not unique, and for general complex numbers there is no particular way to make a standard choice. For positive reals, however, we can define x^n to be the unique *positive* q'th root of x^p. One verifies that this choice is well defined, for if $a/b = p/q$, then $aq = bp$, so that if $y^b = x^a$, then $y^{aq} = y^{bp} = x^{ap}$. By taking unique positive roots, we conclude $y^q = x^p$. If

we want the powers of a real number to be real, we are effectively forced into this path, because then $x^{p/q} = (x^{p/2q})^2 \geqslant 0$.

Really, we have not applied any real analysis yet, other than the continuity and monotonicity of n'th roots. Without more analysis, our attempt to generalize the power function stops. Stated more elegantly, we desire, for each $x > 0$, to find a real-valued continuous function f, for which

$$f(a+b) = f(a)f(b) \quad f(1) = x$$

f must necessarily be positive, for $f(x) = f(x/2)^2 \geqslant 0$. This shows why our programme fails for non-positive numbers. If g also satisfies these formulae, then $g(1) = f(a)$, then $g(m/n) = f(a)^{m/n} = f(am/n)$, and by continuity, letting $m/n \to b$ for any b, we find $g(b) = f(ab)$, so that the product formula automatically holds.

One automatically verifies that f must necessarily be unique in the formula, for if $f(1) = g(1)$, and f and g satisfy the formulae, then one verifies that f and g agree on all rational values, implying by continuity that $f = g$ everywhere. Thus we need only find one formula that works for each positive x, and we shall define $x^t = f(t)$.

For any complex z, we define the exponential function

$$\exp(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

Since

$$\sqrt[n]{\frac{|z|^n}{n!}} = \frac{|z|}{\sqrt[n]{n!}} \leqslant \frac{|z|}{\sqrt{n}} \to 0$$

the series converges absolutely for all z, and therefore converges uniformly on any bounded subset. This implies the exponential function is continuous. Since

$$\frac{d}{dz} \left(\sum_{k=0}^{N} \frac{z^k}{k!} \right) = \sum_{k=0}^{N-1} \frac{z^k}{k!}$$

which converges absolutely to the exponential function, we have the incredibly useful formula $\exp'(z) = \exp(z)$. What's more, the absolute convergence of the series implies the manipulation

$$\sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{z^k w^j}{k!j!} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{z^k w^{n-k}}{k!(n-k)!} = \sum_{n=0}^{\infty} \frac{(z+w)^n}{n!}$$

8

so $\exp(z+w) = \exp(z)\exp(w)$. Thus we have found a function f, for $f(1) = \exp(0)$, which we define to be the number e.

Theorem 2.1.

(a) $\exp(z) \neq z$ for any complex z.

(b) $\exp(t) > 0$ for all $t \in \mathbf{R}$.

(c) $\exp$ is monotonically increasing on the real axis.

(d) $\exp(t) \to \infty$ as $t \to \infty$, and $\exp(t) \to 0$ for $t \to -\infty$.

Proof. (a) is proved by noticing

$$\exp(z)\exp(-z) = \exp(0) = 1$$

What's more, $\exp(\mathbf{R})$ is a connected subset of $\mathbf{R}$ containing 1 but not 0, so $\exp(\mathbf{R}) \subset (0, \infty)$, proving (b). Then (c) is shown, since $\exp' = \exp > 0$. We have the very weak estimate $\exp(t) > 1 + t$ for $t > 0$, so that $\exp(t) \to \infty$ as $t \to \infty$ and therefore (since $\exp(-t) = \exp(t)^{-1}$) that $\exp(t) \to 0$ as $t \to -\infty$. We conclude that the exponential function provides a homeomorphism between $\mathbf{R}$ and $(0, \infty)$. $\square$

This finalizes are discussion of powers, for if x is any positive number, then $x = e^y$ for some unique $y \in \mathbf{R}$ (which we define to be $\log x$, called the logarithm of x), and the function $f(a) = e^{ay}$ satisfies the properties which define an extension of the power function. For negative x, there is no natural choice of y for which $e^y = x$, so that general powers cannot be defined. Nonetheless, we will define $e^z = \exp(z)$, for any complex number z, and this is a continuous function satisfying the desired properties.

Let us switch to studying the miraculous properties of the function e^{it}, for $t \in \mathbf{R}$.

Theorem 2.2.

(a) $|e^{it}| = 1$ for all t.

(b) There exists a number π such that $e^{i(\pi/2)} = i$, and any number satisfying $e^x = 1$ is an integer multiple of $2\pi i$.

(c) The map $t \mapsto e^{it}$ is surjective onto the set $\mathbf{T}$ of complex numbers with absolute value 1, and the exponential map is surjective onto $\mathbf{C} - \{0\}$.

9

Proof. Note that

$$\overline{e^{it}} = \sum_{k=0}^{\infty} \frac{(-it)^k}{k!} = e^{-it}$$

so that

$$|e^{it}|^2 = e^{it}\overline{e^{it}} = e^{it}e^{-it} = 1$$

which is enough to prove (a).

Finding π requires a careful exploration of the power series of e^{it}. Notice it is alternating in each axis of the plane, which enables us to easily get precise estimates on the values of the function. Take the projections of the power series of the exponential onto the real and imaginary axis, obtaining the expansions

$$\mathfrak{Re}(e^{ix}) = \sum_{k=0}^{\infty}(-1)^k \frac{x^{2k}}{(2k)!} \quad \mathfrak{Im}(e^{ix}) = x\sum_{k=0}^{\infty}(-1)^k \frac{x^{2k}}{(2k+1)!}$$

For $x \in (0,2]$ the absolute values of the coeffients begin to decrease for $k \geqslant 1$, so we obtain the bounds

$$\mathfrak{Re}(e^{ix}) \leqslant 1 - x^2/2 + x^4/24 \qquad x(1 - x^2/6) \leqslant \mathfrak{Im}(e^{ix})$$

This implies that $0 < \mathfrak{Im}(e^{ix})$, and $\mathfrak{Re}(e^{2i}) \leqslant 1 - 2 + 2/3 = -1/3$, so that $e^{iy} = i$ for some $y \in (0,2)$. There is only one such solution in this region, because here

$$\frac{d}{dx}\mathfrak{Re}(e^{ix}) = -\mathfrak{Im}(e^{ix}) < 0$$

We define π to be the smallest positive number for which $e^{i\pi} = -1$. It is easy to see $\pi = 2y$, because if $e^{iz} = -1$, then $e^{i(z/2)} = \pm i$, and if $0 \leqslant z/2 \leqslant y$, we must in fact have $e^{i(z/2)} = i$, hence $z/2 = y$. It is also simple to show that $e^{ix} = 1$ holds if and only if x is a multiple of $2\pi i$. First, the set of all such x must lie in $i\mathbf{R}$, because

$$|e^{a+ib}| = e^a|e^{ib}| = e^a$$

and so $|e^{a+ib}| = 1$ if and only if $e^a = 1$, so $a = 0$. Secondly, note that the set of all x for which $e^{ix} = 1$ must form a discrete subgroup G of $\mathbf{R}$, because there is a neighbourhood of the origin such that $e^{ix} \neq 1$ unless $x = 0$. This implies that $G = y\mathbf{Z}$ for the smallest positive $y\mathbf{R}$ satisfying $e^{iy} = 1$. We

must have $e^{i(y/2)} = -1$, hence $y/2 \geqslant \pi$, and so $y \geqslant 2\pi$, and we must in fact have equality since $e^{2\pi i} = 1$.

Now we prove that the map e^{ix} is surjective onto S^1. First, consider $w = x + iy$, with $x, y \geqslant 0$. Then $x \leqslant 1$, and because $\mathfrak{Re}(e^0) = 1$, $\mathfrak{Re}(e^{i\pi/2}) = 0$, there is some $t \in (0, \pi/2)$ for which $\mathfrak{Re}(e^{it}) = x$. Since $\mathrm{Im}(e^{it}) \geqslant 0$, we have

$$\mathrm{Im}(e^{it}) = \sqrt{1 - \mathfrak{Re}(e^{it})^2} = \sqrt{1 - x^2} = y$$

so $e^{it} = w$. If $w = x + iy$ with $x \geqslant 0$, $y < 0$, then $iw = ix - y$, and $x, -y \geqslant 0$, so there is t for which $e^{it} = iw$, and then $e^{i(t-\pi/2)} = w$. The other cases can be shown by employing the same technique. Since any non-zero complex number z can be written aw, where $a > 0$ and $|w| = 1$, there are t, t' such that $a = e^t$, $w = e^{it'}$, and then $z = e^{t+it'}$. $\qquad\qquad\square$

Thus exp is a surjective homomorphism from the additive group of complex numbers to the multiplicative group, with kernel $2\pi i \mathbf{Z}$. If we look at the topological structure of $\mathbf{C}/2\pi i \mathbf{Z}$, we see the shape forms an infinitely long cylinder, which is exactly the way we think of complex multiplication – the imaginary part of a number forms a loop around the origin, and the real part scales the number radially outward from the origin, like a cylinder.

Chapter 3

Convex Functions

Convex functions occur abound in mathematics, and are normaly the 'feasible' objects of study in hard analysis – where we want to obtain explicit inequalities rather than 'limit' arguments that give no guarantee on convergence rates. Geometrically, a function is convex if the graph of the function lies below the straight line of any two points on the graph. Analytically, we can express a convex function f as one for which

$$f((1-t)a + tb) < (1-t)f(a) + tf(b)$$

for $0 < t < 1$.

Convex functions are fairly pleasant objects to study, for they may be upper bounds on an interval (a,b) once we know the definition of the function at a and b. The contrapositive also says that if we can lower bound a convex function at a point x between a and b, then we can obtain lower bounds at b once we know the value of the function at a. Here's an example application.

Theorem 3.1. *A convex function defined on an open interval is continuous.*

Proof. Let $f : (a,b) \to \mathbf{R}$ be convex, and fix $a < c < b$. Since f is upper bound by a line passing through c, f is upper semicontinuous. But for sufficiently small t, we find

$$f(c) \leqslant \frac{f(c+t) + f(c-t)}{2}$$

If $f(c+t) \leqslant f(c) - \varepsilon$, then $f(c-t) \geqslant f(c) + \varepsilon$, or else

$$f(c) < \frac{f(c) - \varepsilon + f(c) + \varepsilon}{2} = f(c)$$

Since we can choose t small enough that $f(c - t) < f(c) + \varepsilon$ for any ε, we must eventually have $f(c + t) > f(c) - \varepsilon$, so that f is also lower semicontinuous. $\square$

The definition of convexity can be expressed as the fact that for $a < b$,

$$f(x) < \frac{f(b) - f(a)}{b - a}(x - a) + f(a)$$

The two definitions are realized equivalent after a coordinate change, for if $x = (1 - t)a + tb$, then $t = (x - a)/(b - a)$. If we rearrange this equation, we find

$$\frac{f(x) - f(a)}{x - a} < \frac{f(b) - f(a)}{b - a}$$

So that the approximation of the tangent line at a is a decreasing function of the approximation point. In this form, the next theorem is essentially trivial.

Theorem 3.2. *If $f : (a, b) \to \mathbf{R}$ is convex, then f' is increasing where defined.*

Proof. Let $x < y < z$, and suppose f is differentiable at x and z. Applying the last definition of inequality twice, we find that

$$f'(x) < \frac{f(y) - f(x)}{y - x} < \frac{f(z) - f(x)}{z - x} < \frac{f(z) - f(y)}{z - y} < f'(z)$$

so f' is increasing. $\square$

Thus a convex differentiable function is either increasing, decreasing, or decreasing to the left of some point and increasing to the right.

It follows that if f is twice differentiable and convex, then $f'' > 0$. The converse is a fact implicitly used in first year calculus courses, but intuitively obvious. It isn't too hard to prove either, with the power of analysis at our hand. Consider the next partial result, which will give us the full result when we apply a Rolle's theorem like argument (this kind of method can always be applied when we are discussing inequalities involving tangent lines, since these inequalities are easy to rotate so that the tangent lines are straight).

Theorem 3.3. *Suppose f is differentiable, and f' is increasing. If $a < b$, and $f(a) = f(b)$, then $f(x) < f(a) = f(b)$ for $a < x < b$.*

13

Proof. We may apply the mean value theorem to find $a < y < b$ for which $f'(y) = 0$. Since f' is increasing, $f' < 0$ on (a, y), and $f' > 0$ on (y, b). But this essentially proves the inequality, because it implies that f decreases to a minimum point, and then increases. The maximum point of f therefore occurs on the boundary. □

Corollary 3.4. *If f' is increasing, then f is convex.*

Proof. Fix $a < x < b$, and consider the function

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a)$$

Then g' is just f' plus a constant, so g' is also increasing. Since $g(a) = f(a) = g(b)$, the last theorem implies that

$$f(x) - \frac{f(b) - f(a)}{b - a}(x - a) = g(x) < f(a)$$

which we can rewrite as

$$\frac{f(x) - f(a)}{x - a} < \frac{f(b) - f(a)}{b - a}$$

so we have convexity! □

This makes it easy to find a plethora of convex functions. Since

$$\frac{d^2 x^n}{dx^2} = n(n - 1)x^{n-2}$$

we find the map $f(x) = x^n$ is convex on $[0, \infty)$, and on $(-\infty, \infty)$ if n is positive and even. The map $f(x) = e^x$ is positive, and since $f''(x) = f(x) > 0$, the exponential function is convex. If we look at the graphs of these functions, we see that the tangent line lower bounds the functions on the whole domain. This is not hard to prove in general.

Theorem 3.5. *If f is convex on (a, b) and differentiable at y, then for $x \neq y$*

$$f(x) > f(y) + f'(y)(x - y)$$

14

Proof. For $x < y < z$,

$$\frac{f(x) - f(y)}{x - y} < \frac{f(z) - f(y)}{z - y}$$

Letting $x \to y$, we find

$$f'(y) < \frac{f(z) - f(y)}{z - y}$$

so $f(y) + f'(y)(z - y) < f(z)$, and by letting $z \to y$, we find

$$\frac{f(x) - f(y)}{x - y} < f'(y)$$

so

$$f(x) > f(y) + f'(y)(x - y)$$

and we have obtained lower bounds both to the left and right of y. $\qquad\square$

The converse is also true

Theorem 3.6. *If the graph of a differentiable function lies above the tangent line at a point, then the function is convex.*

Proof. s $\qquad\square$

We can also consider **concave functions**, which are lower bounded by the lines between points. Any concave function can be obtained from a convex function by taking the negative, and conversely any convex function can be obtained from a concave function, so they are essentially the same class of functions, and by negating f in the theorems above we obtain a family of results for concave functions. For instance, the graph of a concave function lies below any tangent line, twice differentiable concave functions are precisely those for which $f'' < 0$, and f' is decreasing where defined even when the function is not twice differentiable.

If f is a twice continuously differentiable function, then at almost every point on the domain it is either locally convex or locally concave. If $f''(x) > 0$, then $f'' > 0$ in a neighbourhood around x, so f is locally convex around x, and if $f''(x) < 0$, then $f'' < 0$ in a neighbourhood around x, so f is locally concave. The only problems are inflection points, when $f''(x) = 0$, but these can normally be locally analyzed (they are normally where the graph twists from convex to concave, or vice versa). For instance, x^3 is convex on $(0, \infty)$, concave on $(-\infty, 0)$, and has an inflection point at 0.

Theorem 3.7. *If the tangent lines of a differentiable function $f : (a,b) \to \mathbf{R}$ intersect the graph of f only once, then f is either convex or concave.*

Proof. We first prove that no straight line can intersect f more than three times. Indeed, if a straight line intersection $a < b < c$, then we would find

$$\frac{f(b) - f(a)}{b - a} = \frac{f(c) - f(a)}{c - a}$$

If we consider

$$g(x) = \frac{f(x) - f(a)}{x - a}$$

Then $g(b) = g(c)$, so there is $b < y < c$ with $g'(y) = 0$, so

$$f'(y)(y - a) = [f(y) - f(a)]$$

and this can be rephrased as $f(y) + f'(y)(a - y) = f(a)$, so that the tangent line at y intersect the graph of f at more than one place.

Now we consider the function

$$g(x,y,t) = \frac{f((1-t)x + ty) - f(x)}{ty - tx} - \frac{f(y) - f(x)}{y - x}$$

defined for $x < y$, $0 < t < 1$. By the argument above, $g(x,y,t) \neq 0$ on the domain, so by continuity $g(x,y,t) > 0$ for all values, or $g(x,y,t) < 0$, so f is either convex or concave. $\square$

Convexity behaves slightly well under composition, but not as well as we'd like it too.

Theorem 3.8. *If f is increasing and convex, and g is also convex, then $f \circ g$ is convex.*

Proof. We prove by calculation, finding that $g(tx + (1-t)y) \leqslant tg(x) + (1-t)g(y)$, so, since f is increasing,

$$(f \circ g)(tx + (1-t)y) \leqslant f(tg(x) + (1-t)g(y)) \leqslant t(f \circ g)(x) + (1-t)(f \circ g)(y)$$

If f is not increasing, then we cannot guarantee $f \circ g$ is convex. Indeed, if $f = e^{-x}$, then $f'' = e^{-x} > 0$, but $g = f \circ f = e^{e^{-x}}$, so $g' = -e^{-x}e^{e^{-x}}$, and

$$g'' = e^{-x}e^{e^{-x}} + e^{-2x}e^{e^{-x}} = e^{-x}e^{e^{-x}}(1 + e^{-x})$$

which is not necessarily positive for x small enough. $\square$

16

The main inequality for convex functions is very important in almost any part of hard analysis. First, consider $a_1,\ldots,a_n > 0$ with $\sum a_i = 1$. Given $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$, which aren't all equal, we have the trivial interpolation inequality $x_1 < \sum_{i=1}^{n} a_i x_i < x_n$. If $t = \sum_{i=1}^{n-1} a_i = 1 - a_n$, then we also have

$$tx_1 < \sum_{i=1}^{n-1} a_i x_i < tx_n$$

obtained by subtracting $a_n x_n$ from the inequality above. With these trivial observations, we can prove a nontrivial inequality.

Theorem 3.9. *If f is convex, and defined on an interval containing all x_i, then $\sum a_i = 1$ and $a_i > 0$, then $f(\sum a_i x_i) < \sum a_i f(x_i)$*

Proof. Using the basic definition of convexity, we can complete the proof by induction. We find

$$f\left(\sum_{i=1}^{n} a_i x_i\right) = f\left((1-a_n)\sum_{i=1}^{n-1} \frac{a_i}{1-a_n} x_i + a_n x_n\right) < (1-a_n)f\left(\sum_{i=1}^{n-1} \frac{a_i}{1-a_n} x_i\right) + a_n f(x_n)$$

By induction, we find

$$f\left(\sum_{i=1}^{n-1} \frac{a_i}{1-a_n} x_i\right) < \sum \frac{a_i}{1-a_n} f(x_i)$$

and this completes the proof by substitution. The previous arguments just show that the manipulations are justified since the values considered lie in the domain of f. $\square$

As an example of this theorem, the exponential is a convex function, so that for any $x_1,\ldots,x_n$,

$$\sqrt[n]{e^{x_1} e^{x_2} \ldots e^{x_n}} = e^{\sum x_i/n} < \frac{1}{n}\sum e^{x_i}$$

If $y_i = e^{x_i}$, we discover the arithmetic geometric mean inequality

$$\sqrt[n]{y_1 \ldots y_n} < \frac{y_1 + \cdots + y_n}{n}$$

One of the classical inequalities in analysis.

This theorem can be generalized to a very important theorem in probability theory – if X is a random variable, and f is convex, then $f(\mathbf{E}X) \leqslant \mathbf{E}[f(X)]$. To see this, note that if X is a simple random variable of the form $\sum a_i \mathbf{I}(X \in E_i)$, then applying the discrete Jensen's inequality we find

$$f(\mathbf{E}X) = f\left(\sum \mathbf{P}(X \in E_i)a_i\right) \leqslant \sum \mathbf{P}(X \in E_i)f(a_i) = \mathbf{E}[f(X)]$$

If X is a general random variable, then X can be expressed as the monotone limit of simple random variables $X_1 \leqslant X_2 \leqslant \cdots \to X$, and then the monotone convergence theorem guarantees that $\mathbf{E}X_i \to \mathbf{E}X$, so that by continuity

$$f(\mathbf{E}X) = \lim_{n\to\infty} f(\mathbf{E}X_n) \leqslant \liminf_{n\to\infty} \mathbf{E}[f(X_n)]$$

The proof would be complete, but it's hard to limit the right hand side, so we appeal to a more abstract proof.

Theorem 3.10. *If f is convex on $-\infty \leqslant a < b \leqslant \infty$, $f(\mathbf{E}X) \leqslant \mathbf{E}[f(X)]$.*

Proof. Let $y = \mathbf{E}X$, then $a \leqslant y \leqslant b$. Let

$$\alpha = \sup_{a \leqslant x \leqslant y \leqslant b} \frac{f(y) - f(x)}{y - x}$$

Then $\alpha < \infty$ by convexity, and for any $a \leqslant x \leqslant y$, $f(x) \geqslant f(y) + \alpha(x - y)$. But then by integrating boths dies we find

$$\mathbf{E}[f(X)] \geqslant f(y) + \alpha(\mathbf{E}[X] - y) = f(y) = f(\mathbf{E}X)$$

$\square$