MC2 Reflection 1

Gabe Orosan-Weine

May 2023

1 Limits

To define limits, we first introduce the notion of a limit point.

Definition 1.1 (Limit point)**.** The point c is a limit point of the set A if for all $\varepsilon > 0$, the interval $(c - \varepsilon, c + \varepsilon)$ contains some $a \in A$ with $a \neq c$.

Limit points are points which have neighborhoods of points surrounding them. The limit of a single variable function can be taken as that variable approaches a constant, or positive or negative infinity, and can be equal to a constant, positive or negative infinity, or fail to exist. The limit of a function $f(x)$ at a limit point c is defined based on its behavior at these points.

Definition 1.2 (Limit approaching a point)**.** Let $f : A \to \mathbb{R}$, and let c be a limit point of A. $\lim_{x\to c} f(x) = L$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that $|x - c| < \delta$ implies $|f(x) - L| < \varepsilon$.

The limit of $f(x)$ is equal to a constant L if, by taking x-values sufficiently close to c, it is possible to get $f(x)$ arbitrarily close to $f(L)$. For example, $\lim_{x\to 3} 2x - 5 = 1$ because for every $\varepsilon > 0$, there exists a $\delta = \varepsilon/2$ such that $0 < |x - 3| < \delta \implies |2x - 5 - 1| = 2|x - 3| < 2\frac{\varepsilon}{2} = \varepsilon$.

Definition 1.3 (Limit approaching infinity)**.** Let $f : A \to \mathbb{R}$. $\lim_{x\to\infty} f(x) = L$ if for every $\varepsilon > 0$, there exists $n \in \mathbb{R}$ such that $x > n$ implies $|f(x) - L| < \varepsilon$.

In this case, the limit as x approaches infinity is a value L that $f(x)$ can be made arbitrarily close to by taking sufficiently large values of x. In both cases the limit fails to exist if the condition cannot be satisfied.

The point definition of limits gives us a direct way to define continuity.

Definition 1.4 (Continuity)**.** A function $f : A \to \mathbb{R}$ is continuous at a point $c \in A$ if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $|x - c| < \delta$ implies $|f(x) - f(c)| < \varepsilon$.

Notice that we do not require c to be a limit point but require that it is in the domain, A. We require that the value of the function at c is equal to its limit as x approaches c, asserting that the limit is equal to $f(c)$ and dropping the condition that $|x - c| > 0$. If c is a limit point of A, then this definition is equivalent to saying that f is continuous at c if

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

A continuous function is a function that is continuous at all points in its domain.

Defining continuity gives us a way to put some algebra behind the notion that a continuous function is one you can draw without picking up your pencil, and can be adapted to take the limit of multiple variables by using vectors and euclidean distance.

Definition 1.5 (Derivative)**.** Let $f : A \to \mathbb{R}$ be a function defined on an interval A. The derivative of f at $x \in A$ is defined as

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

The derivative tries to capture the rate of change of a function around a point. This formulation allows us to derive the power rule for positive n. In order to do so, we start with a lemma:

Lemma 1.1. Let c be a constant, and $n \geq 1$. $\lim_{h \to 0} h^n c = 0$.

Proof. Assume c is a constant. Given $\varepsilon > 0$, if $c = 0$ then all h values give $|hc - 0| = |0 - 0| = 0 < \varepsilon$. Otherwise if $c \neq 0$, pick $\delta = \min(\{\varepsilon/|c|, 1\})$. Then $|h - 0| < \delta$ implies

$$-\frac{\varepsilon}{|c|} < h - 0 < \frac{\varepsilon}{|c|}.$$

Multiplying everything by $|c|$, we have

$$-\varepsilon < h|c| - 0 < \varepsilon$$

$$|hc| < \varepsilon.$$

Additionally, $\delta \leq 1$ so $|h| < 1$, which implies $|h^n| \leq |h|$. Therefore we have

$$|h^n c - 0| \leq |hc| < \varepsilon.$$

$\square$

Theorem 1.2 (Power Rule)**.** Let $f(x) = x^n$ with $n \in \mathbb{N}$. $f'(x) = nx^{n-1}$.

Proof. Assume $f(x) = x^n$ with $n \in \mathbb{N}$.

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{(x+h)^n - x^n}{h}.$$

By the binomial theorem,

$$(x+h)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} h^i.$$

We have a single term which contains solely x^n, which is then subtracted, and a single term with a power of 1 for h so we can simplify as follows.

$$\lim_{h\to 0} \frac{\sum_{i=0}^{n} \binom{n}{i} x^{n-i} h^i - x^n}{h} = \lim_{h\to 0} \frac{x^n - x^n}{h} + \frac{nx^{n-1}h}{h} + \frac{\sum_{i=2}^{n} \binom{n}{i} x^{n-i} h^i}{h}.$$

The first term reduces to $0/h$ which is equal to 0, and the last group of terms all have exponents of h greater than 1 after accounting for the denominator, so each of their limits is 0. Thus we are left with the single term in the middle being nonzero.

$$f'(x) = \lim_{h\to 0} nx^{n-1} = nx^{n-1}.$$

□

Limits are surprisingly useful in understanding transcendental numbers through convergent series.

Definition 1.6 (Powers of Euler's Number)**.** We define Euler's Number e raised to the power x as

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Equivalently,

$$e^x = \lim_{n\to\infty} s_n.$$

where

$$s_n = \sum_{n=0}^{n} \frac{x^n}{n!}.$$

From this definition, we can take the derivative of $f(x) = e^x$ by taking the derivative of each term in the sum.

Theorem 1.3. Let $f(x) = e^x$. $f'(x) = e^x$.

Proof.

$$f(x) = \lim_{n\to\infty} \sum_{n=0}^{n} \frac{x^n}{n!}.$$

To take the derivative of $f(x)$, we can evaluate the derivative of each term in the sequence, then add them afterwards. 0! is equal to 1 so the derivative of the first term in the sequence is 0. Since $n!$ is a constant, the derivative of $\frac{x^n}{n!}$ for $n \geq 1$ is equal to $\frac{nx^{n-1}}{n!} = \frac{x^{n-1}}{(n-1)!}$. So we have

$$f'(x) = \lim_{n\to\infty} \sum_{n=1}^{n} \frac{x^{n-1}}{(n-1)!} = \lim_{n\to\infty} \sum_{n=0}^{n} \frac{x^n}{(n)!} = e^x.$$

□

This demonstrates that the notion of limits and derivatives can give us insight into deep properties of numbers and functions.

2 Gradients

Definition 2.1 (Gradient)**.** The gradient of a multivariable function $f(x_1, x_2, ..., x_n)$ is defined as

$$\nabla f(x_1, x_2, ..., x_n) = \langle f_{x_1}, f_{x_2}, ... f_{x_n} \rangle$$

where f_x is the partial derivative of f with respect to x.

The gradient of a multivariable function, containing the partial derivative of the function with respect to each of its input variables, gives us useful insight into the rate of change of the function and has uses almost any time you are trying to characterize, approximate, or optimize a function of more than one variable.

Analogous to how the derivative lets us find a tangent line in single variable functions, we can use the gradient to construct a tangent plane to a surface described by a multivariable function. First, we define a plane and tangent line as follows.

Definition 2.2 (Plane)**.** A plane is determined by a point (x_0, y_0, z_0) and a normal vector $\vec{V} = \langle a, b, c \rangle$. Points in the plane are points (x, y, z) where V is orthogonal to the vector $\vec{P} - \vec{P_0}$ meaning $(\vec{P} - \vec{P_0}) \cdot \vec{V} = 0$, where $\vec{P_0} = \langle x_0, y_0, z_0 \rangle$ and $\vec{P} = \langle x, y, z \rangle$. Thus an equation for a plane with the point (x_0, y_0, z_0) and normal vector $\vec{V} = \langle a, b, c \rangle$ is given by

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

Definition 2.3 (Tangent Line)**.** A tangent line to a curve $y = f(x)$ at a point (x_0, y_0) is given by $y - y_0 = f'(x_0)(x - x_0)$.

Notice that the point (x_0, y_0) satisfies this equation, and the derivative with respect to x is equal to $f'(x_0)$.

We define a tangent plane to a surface S at point (x_0, y_0, z_0) as a plane passing through (x_0, y_0, z_0) which contains the tangent lines of all the curves on S passing through (x_0, y_0, z_0). If S is given by $z = f(x, y)$, one way to find a tangent plane is to set the dot product $\nabla f(x_0, y_0, z_0) \cdot (\vec{P} - \vec{P_0}) = 0$, where $\vec{P_0} = \langle x_0, y_0, z_0 \rangle$ and $\vec{P} = \langle x, y, z \rangle$.

Theorem 2.1 (Tangent Plane)**.** Let the surface S be given by $f(x, y, z) = k$ and let (x_0, y_0, z_0) be a point on S. If f has continuous first partial derivatives, then the plane determined by the normal vector ∇f and point (x_0, y_0, z_0) is a tangent plane to S.

Proof. We will show that the plane T given by

$$\nabla f(x_0, y_0, z_0) \cdot (\vec{P} - \vec{P_0}) = 0$$

where $\vec{P_0} = \langle x_0, y_0, z_0 \rangle$ and $\vec{P} = \langle x, y, z \rangle$, contains all tangent lines of curves on S passing through (x_0, y_0, z_0). Let C be a curve on S passing through

(x_0, y_0, z_0) described by a continuous vector function $\vec{r}(t) = \langle x(t), y(t), z(t)\rangle$ where $\vec{r}(t_0) = (x_0, y_0, z_0)$. Since C is on S,

$$f(x(t), y(t), z(t) = k.$$

Using the chain rule, we get

$$\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = 0$$

which can be written as

$$\nabla f \cdot \vec{r}\,'(t) = 0.$$

Since $\vec{r}(t_0) = \langle x_0, y_0, z_0\rangle$ we have

$$\nabla f(x_0, y_0, z_0) \cdot \vec{r}\,'(t_0) = 0.$$

The gradient vector at (x_0, y_0, z_0) is orthogonal to the tangent vector $\vec{r}\,'(t_0)$ to any curve C on S passing through (x_0, y_0, z_0). Thus, the plane determined by the normal vector $\nabla f(x_0, y_0, z_0)$ and the point (x_0, y_0, z_0), given by

$$\nabla f(x_0, y_0, z_0) \cdot (\vec{P} - \vec{P}_0) = 0$$

constitutes a tangent plane to S at (x_0, y_0, z_0). $\square$

In the special case that S is given by $z = g(x, y)$, we have $f(x, y, z) = g(x, y) - z$ so $f_z = -1$ and our equation for the tangent plane becomes

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Moving z_0 to the right-hand side, we have

$$z = z_0 + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Just like tangent lines at a point let us compute linear approximations for single variable functions, tangent planes can be used to make estimates for multivariable functions. To find an approximation for $z = g(x, y)$ above, we use the tangent plane at $(a, b, g(a, b))$ to get

$$g(x, y) \approx g(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

Gradients are also useful because the maximum rate of change of a function is in the direction of the gradient.

Definition 2.4 (Directional Derivative)**.** The directional derivative of f at (x_0, y_0, z_0) in the direction of a unit vector $\vec{u} = \langle a, b, c\rangle$ is

$$D_u f(x_0, y_0, z_0) = \lim_{h\to 0} \frac{f(x_0 + ha, y_0 + hb, z_0 + hc) - f(x_0, y_0, z_0)}{h}.$$

Theorem 2.2. If $f(x, y, z)$ is differentiable then f has a directional derivative in direction of any unit vector $\vec{u} = \langle a, b, c\rangle$ and

$$D_u f(x, y, z) = \nabla f(x, y, z) \cdot \vec{u} = f_x(x, y, z)a + f_y(x, y, z)b + f_z(x, y, z)c.$$

Proof. Let $g(h) = f(x_0 + ha, y_0 + hb, z_0 + hc)$. By the definition of a derivative,

$$g'(0) = \lim_{h\to 0} \frac{g(h) - g(0)}{h} = \lim_{h\to 0} \frac{f(x_0 + ha, y_0 + hb, z_0 + hc) - f(x_0, y_0, z_0)}{h} = D_u f(x_0, y_0, z_0).$$

By setting $x = x_0 + ah, y = y_0 + bh$, and $z = z_0 + ch$, we can use the chain rule to write

$$g'(h) = \frac{\partial f}{\partial x}\frac{dx}{dh} + \frac{\partial f}{\partial y}\frac{dy}{dh} + \frac{\partial f}{\partial z}\frac{dz}{dh} = f_x(x, y, z)a + f_y(x, y, z)b + f_z(x, y, z)c.$$

Plugging in $h = 0$ so $x = x_0, y = y_0$, and $z = z_0$, we find

$$g'(0) = f_x(x_0, y_0, z_0)a + f_y(x_0, y_0, z_0)b + f_z(x_0, y_0, z_0)c.$$

Setting the two results equal to each other, we have

$$D_u f(x_0, y_0, z_0) = f_x(x_0, y_0, z_0)a + f_y(x_0, y_0, z_0)b + f_z(x_0, y_0, z_0)c.$$

□

Theorem 2.3. Suppose $f(x, y, z)$ is differentiable. The greatest value of the directional derivative $D_u f(x, y, z)$ is $|\nabla f(x, y, z)|$ which is attained when $\vec{u}$ is in the direction of the gradient $\nabla f(x, y, z)$.

Proof. Since $\vec{u}$ is a unit vector,

$$D_u f = \nabla f \cdot \vec{u} = |\nabla f||\vec{u}| \cos\theta = |\nabla f| \cos\theta$$

where θ is the angle between ∇f and $\vec{u}$. Because $\cos\theta \leq 1$ for all θ,

$$D_u f = |\nabla f| \cos\theta \leq |\nabla f|.$$

When $\vec{u}$ is in the direction of ∇f, $\theta = 0$ so $\cos\theta = 1$ and $D_u f = |\nabla f|$. □

This result is part of what makes the gradient so useful in optimization problems. One prominent example of this is in machine learning where the goal of the most widely used learning algorithm, backpropagation, is to find parameters which minimize a loss function for a given input. This is done by gradient descent - subtracting a multiple of the loss function's gradient from the parameters. The biggest change in the loss function with the smallest change in parameters is obtained when you go in the direction of the gradient.

In summary, gradients are very useful for understanding how multivariate functions change with respect to their inputs, which has applications in approximations and optimization among many other types of problems.