3. Conversations with $\mathbb{R}$

In previous two chapters, we learned about the alphabet and words used in mathematical language, which is $\mathbb{R}$. Now we are ready to have conversations with these words. Since math words $\mathbb{R}$ is closed under the operations addition and multiplication, although alphabets generate words and words generate sentences, they mean the same in math.
In this chapter, we will mainly be talking about 1 input 1 response conversations, meaning that for every spoken input, there will be only one response. With this idea, we can start to analyze lots of interesting conversations, their emotions and information presented in these conversations.

3.1 Functions

3.1.1 The Basics

Let us start by defining general conversation between two person using two alphabets (which are sets). Let us keep in mind that when we are dealing with specific mathematical structures that are closed under some operation, alphabets have the same meaning as words and sentences.

Definition 3.1.1 A **function** f between sets X and Y (conversation between two person using two alphabets) is a communication log that corresponds each $x \in X$ with an element $y \in Y$. We say that "x maps to y by f", and we often denote $f : X \to Y, x \mapsto y$ and $y = f(x)$. One calls y the **image** of x and x the **preimage** of y. We call X the **domain** of f and Y the **codomain** of f. We call $f(X) = \{f(x) : x \in X\}$ to be the **image** of f.

■ **Example 3.1 — Functions.** ■

- In the set $X = \{1,2,3\} = Y$, we can have a function $f : X \to Y$ such that $f(1) = 2$, $f(2) = 3$ and $f(3) = 1$. We can also have a function $g : X \to Y$ such that $g(1) = g(2) = g(3) = 1$. However, if we associate $1 \in X$ to both $2 \in Y$ and $3 \in Y$, then we will not get a function. An important thing to note here is that for every $x \in X$ and any function $f : X \to Y$, $f(x)$ is **unique**.
- When $X = Y = \mathbb{R}$, we can get a lot of familiar functions, such as $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ for all $x \in \mathbb{R}$. We can see that $1 \mapsto 1$, $2 \mapsto 4$ and $-3 \mapsto 9$. We can see that the *image* of f is

$\mathbb{R}^+ \cup \{0\}$. Another example can be $g : \mathbb{R} \to \mathbb{R}$ where $g(x) = \sin(x)$ for all $x \in \mathbb{R}$. In this case, $0 \mapsto 0$, $\frac{\pi}{2} \mapsto 1$ and $\frac{3\pi}{2} \mapsto -1$. In this case, the image of g is $[-1,1]$.
- When $X = \mathbb{N}$, $Y = \mathbb{R}$, and if we have any function $f : X \to Y$, then we call the set of images of f, namely $\{f(1), f(2), \ldots\}$ a (real) **sequence**. As an example, if we have $f : \mathbb{N} \to \mathbb{R}$ such that $f(x) = x+2$, then $f(1) = 3, f(2) = 4, f(3) = 5, \ldots$ is a sequence.
- In previous sections, we haven't defined the notion of "norm" rigorously. In fact, a norm N on $\mathbb{R}$ is a function $N : \mathbb{R} \to \mathbb{R}$ such that $x \mapsto |x|$ and satisfying the norm defining properties.

R A very important thing to notice is that in the definition of functions, the person with alphabet X uses *all* letters in the alphabet, while the person with alphabet Y does not necessarily use up all letters in Y, i.e. every $x \in X$ is mapped by f to some $y \in Y$, but not every $y \in Y$ is mapped from some $x \in X$.

From above remark and the definition of functions, we understand that not necessarily all letters in Y are mapped from some letter in X ($f(X)$ may not be equal to Y) and not necessarily each letter in X is mapped differently to some letter in Y. Functions with these type of properties are more desirable and very useful for analysis, so we would like to give them specific terminologies.

Definition 3.1.2 We say a function $f : X \to Y$ is **injective** if for all $x, x' \in X$, $f(x) \neq f(x')$. We say a function $f : X \to Y$ is **surjective** if for all $y \in Y$, there exists $x \in X$ such that $x \mapsto y$, i.e. $f(X) = Y$. We say f is **bijective** if it is both injective and surjective.

■ Example 3.2 — Injectivity, Surjectivity and Bijectivity. ■

- Same as in example 3.1.1, if $X = \{1,2,3\} = Y$, f defined in 3.1.1 is both *injective* and *surjective*, hence *bijective*. However, the g defined in 3.1.1 is not *injective* nor *surjective*. For any set $X = Y = \{1,2,3,\ldots,n\} \subset \mathbb{N}$, the set of bijective functions $f : X \to Y$ is called the set of **permutations** on X, since each function is essentially "permuting" the elements $1,2,\ldots,n$.
- We call a function $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = a_0 + a_1x + \ldots + a_nx^n$ for all $x \in \mathbb{R}$ where $a_i \in \mathbb{R}$ for all i and $n \in \mathbb{N}$ a (real) **polynomial** of degree n. An example would be $f(x) = x^2$, this is a real polynomial of degree 2, we can see that in this case f is not surjective or injective, since $f(1) = f(-1)$ and $f(x) \geq 0$ for all $x \in \mathbb{R}$. However, one can show that $f(x) = x^3$ is *bijective*. (*Hint:* Showing it is surjective is simple, since we can solve $x^3 = a$ for all $a \in \mathbb{R}$. To show it is injective, consider $x^3 - y^3 = 0$ and factor.)

It is a common scenario for us to consider only the conversation involving 1 person saying *specific* "words" and the other person's responses. This natural language intuition gives rise to the "restriction" of a function.

Definition 3.1.3 Given $f : X \to Y$, and $A \subset X$, we define the **restriction of f to A** to be the function $f|_A : A \to Y$ such that $f|_A(a) = f(a)$ for all $a \in A$. It is easy to see that since f is a function, $f|_A$ is also a function.

R $f|_A$ can be seen as the conversation log for usage of only words in A and their responses.

■ Example 3.3 — Function Restriction. ■

- Sometimes non-injective functions can restrict to injective ones. As an example, $f(x) = x^2$ on $\mathbb{R}$ restricts to an injective function $f|_{\mathbb{R}^+} : \mathbb{R}^+ \to \mathbb{R}$.

- By changing the codomain to be the function's image, we get a surjective function. With same example as above, we know that $f(x) = x^2$ is surjective if we consider $f : \mathbb{R} \to \mathbb{R}^+ \cup \{0\}$.

3.1.2 Functions with Group Structure

Let us consider two functions $f : X \to Y$ and $g : X \to Y$ between set X and group Y closed under $\star$. Then, we can define $f \star g$ such that $f \star g : X \to Y$ where $x \mapsto f(x) \star g(x)$. It is easy to see that $f \star g$ is indeed a function.

Intuitively, this means that if we have two separate conversations between two people using the same two alphabets, we can consider the response emotion together for each input. This is what we do all the time, consider everybody's opinion on the same input we provide. Everybody responding needs to use the same alphabet otherwise the combination won't make any sense. In fact, this gives the group operation for functions from X to Y, since multiplying two such functions still gives us a function from X to Y, provided that Y has a group structure.

Similarly, if Y is a (ring) field with another operation $\cdot$, we can also define $f \cdot g : x \mapsto f(x) \cdot g(x)$, and $-f : x \mapsto -f(x)$ and possibly also $f^{-1} : x \mapsto f(x)^{-1}$. We can see that the set of functions from given set X to given set Y has the exact same structure that Y has, with the operations defined as above.

What is truly powerful is that the set of bijective functions $f : X \to X$ for any set X with or without any structure is already a group, equipped with the following multiplication method:

Definition 3.1.4 Given any sets X, Y, Z, and functions $f : X \to Y$, $g : Y \to Z$, we define the **composition of f and g**, read as "g compose f" or "f pull back g" to be $f^*g : X \to Z$ such that $x \mapsto (f^*g)(x) = g(f(x))$.

R The above definition is a bit general, but we can see that for two functions $f, g : X \to X$, we can compose them together using composition defined above. It is important that we keep the codomain of f and domain of g to be the same for composition to make sense. Hence for functions from X to X, we assume them to be bijective for composition to make sense. Many sources use $g \circ f$ to represent g compose f, however it is applying f first, then g. The backward order in the terminology here is why I do not like it. I will use f^*g in all later occurrences.

In fact, with above remark we can see that given any set X, the set of functions from X to itself is **closed** under composition. It is easy to see that this operation is associative by definition.

Theorem 3.1.1

$$(f^*g)^*h = f^*(g^*h)$$

Proof. On the left hand side, we get $x \mapsto g(f(x))$ by f^*g pull back, then $g(f(x)) \mapsto h(g(f(x)))$ by h. On the right hand side, we get $x \mapsto h(g(x))$ by g^*h, then mapsto $h(g(f(x)))$ by f pullback. They are equal. □

Then another question before we conclude that given any set X, the set of bijective functions from X to itself is a group under composition: What is the identity element? In fact, it is easy to see that $f : X \to X$ such that $x \mapsto x$ is the identity element under composition (Try it!). This function is called the **identity function**, denoted id.

In fact, we also need to define the inverse operation of composition.

Definition 3.1.5 Given any bijective function $f: X \to Y$, we define the **inverse function** of f to be $f^{-1}: Y \to X$ such that $f^*(f^{-1}) = id = (f^{-1})^*(f)$, i.e. $f^{-1}: y = f(x) \mapsto x$.

R The reason for this definition requiring f being bijective is that we need every $y \in Y$ to be the image of some $x \in X$, suggesting that f must be *surjective*. Besides, in order for f^{-1} to be a function, we need each $y \in Y$ only has one $x \in X$ such that $f(x) = y$, meaning that f must also be *injective*.
Another thing is that we use f^{-1} both for multiplicative inverse at the beginning of this chapter and for composition inverse in the above definition. To ressolve ambiguity, we *always* use f^{-1} for composition inverse, and $\frac{1}{f}$ for multiplicative inverse, when Y is a group under multiplication.

Theorem 3.1.2 Given any set X, the set of bijective functions from X to itself forms a group under operation composition.

When you have a story, you tell person A, person A retells it to person B. The emotion that person B gets will be the emotion getting from person A telling person A's emotion getting from your version of the story. This is the composition operation! If the response of person A uses all letters in person A's alphabet and you have 1-1 correspondance between your input and person A's response, it is possible to based on person A's response to figure out what you said to person A, we usually use this technique when we are trying to remember what we have said. This is inversion.

■ **Example 3.4 — Composition and Inversion of Functions.** ■

- The set of polynomials on $\mathbb{R}$ of any finite degree (as defined before) is a **ring**. Why? First, it is closed under addition with additive identity being 0 (treated as a 0 degree constant polynomial). The additive inverse is also defined, being negative of each polynomial. Then, it is closed under multiplication, since we get a polynomial when two polynomials multiply, the identity for this operation is the constant polynomial 1. However, it is important to note that polynomials with addition and function composition is *not* a ring. This is because not all polynomials are bijective, i.e. $f(x) = x^2$.
- Given set $X = \{1, 2, 3, \ldots, n\}$, we know that the set of *permutations* is a group from Theorem 3.1.2, since all permutations are bijections. In fact, we call this group S_n for each $n \in \mathbb{N}$. It is a very important group in abstract algebra!

3.1.3 Graphs of Functions

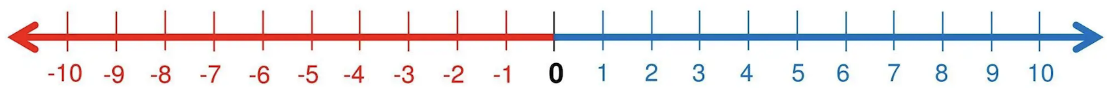
For functions from $\mathbb{R}$ to $\mathbb{R}$, we can actually plot them on a piece of paper. This involves the idea of defining the *graph* of functions, which also leads to one definition of the information of each conversation, which will be covered in future chapters. Let us first define abstractly what it means for the graph of functions.
Before we define the graphs of functions, we will first define a new operation on sets, called the Cartesian Product.

Definition 3.1.6 Given two sets X, Y, we define the **Cartesian Product** of X and Y to be the set $X \times Y = \{(x, y) : x \in X, y \in Y\}$ which is the set of ordered pairs of elements in X and Y.

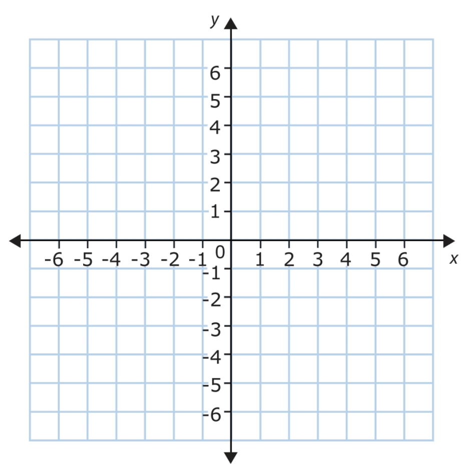
R By *ordered pairs*, we mean that even if $X = Y$, for $a \neq a' \in X = Y$, (a, a') and (a', a) are *two distinct elements* in $X \times Y$.

Definition 3.1.7 Given a function $f : X \to Y$, we define the graph of f, $\text{Graph}(f) : X \to X \times Y$, such that $x \mapsto (x, f(x))$. This is actually a function from X to the Cartesian Product of X and Y. We call the image of this function **the graph of f**.

The intuition behind "graph" is that it is a conversation logging between two person using alphabets X and Y, but instead of logging only the 2nd person's response, we also log the 1st person's input. After defining abstractly what a *graph* of function means, we need to know how to draw the graph. As a starting point, we will define the notion of a "number line." A **number line** is a visual representation of $\mathbb{R}$, where numbers in $\mathbb{R}$ are ordered from left to right, as shown below:

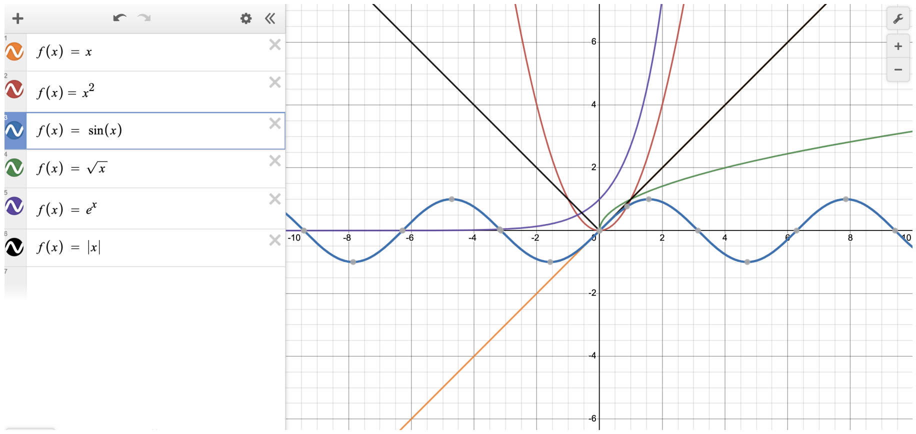


Moving from right to left, the number gets smaller. Moving from left to right the number gets bigger. The red half of the number line above represents *negative* emotions, while the blue half represents *positive* emotions. In order to graph a function $f : \mathbb{R} \to \mathbb{R}$, or even a function $f : X \to Y$ where $X, Y \subset \mathbb{R}$, we first combine two number lines together to make a number plane, like below:

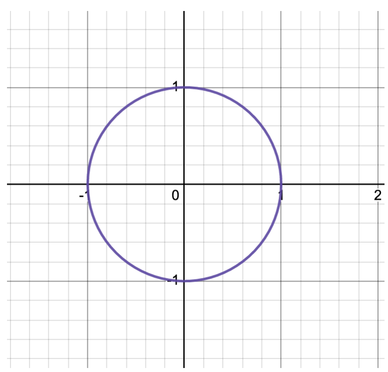


Then, we will label the horizontal axis the x-axis, and the vertical axis y-axis or $f(x)$-axis. Each point on the plane can be labelled by (x, y) where x represents its horizontal position and y represents its vertical position. We call the point $(0,0)$ the **origin**.
In fact, given any function $f : \mathbb{R} \to \mathbb{R}$, each element in the graph of f is a point on the number plane. After we find all the points on the number plane, we connect them together to form the "graph" of function f.

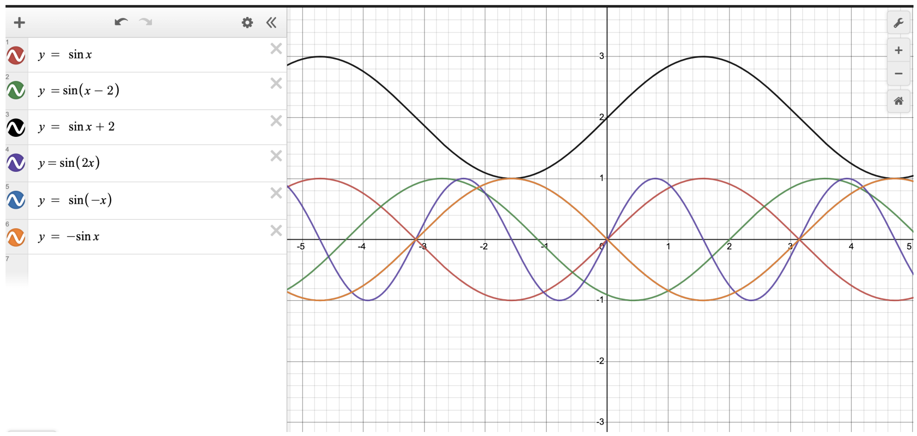
■ **Example 3.5 — Graphs of Functions.** ■



Above is the graphs of several important functions, some of which we haven't seen/defined, but we will in the future. A great tool to use is Desmos, where you can see the graph of all kinds of functions from $\mathbb{R}$ to $\mathbb{R}$. In fact, we can graph all kinds of shapes on the number plane, but is everything representing a function? Not necessarily, let us see an example.



If this is a function, the point 0.5 has two points being its image, which is impossible. It is not hard to realize that for every graph of functions, for every $x \in \mathbb{R}$, if we draw a vertical line through x on the x-axis, this line can only meet the graph at *one point*. In fact, this is a way to tell whether a graph represents a valid function, called **the Vertical Line Test.**
Let us consider the $f(x) = x^2$ graph included above, we know that this function is not *injective*, but can we tell from the graph? Yes, in fact, if we draw a horizontal line through 2 on the y-axis, we can see that it meets the graph at 2 points, meaning that two x values are mapped to the same y value, hence $f(x) = x^2$ is not injective. We can check every horizontal line from each value on the y-axis with the graph to see whether a function is injective or not, this is called the **Horizontal Line Test**.
There are several things you can do to a graph, including translation ($y = \sin(x) + 2$, $y = \sin(x - 2)$), stretching ($y = \sin(3x)$), and reflection $y = \sin(-x)$, $y = -sin(x)$. See below for the effects of these operations:



3.1.4 Countability

In this subsection, we will describe a very important idea called *countability*, meaning whether a set S can be *counted* with natural numbers. By definition of natural numbers, we know that S can be finite or infinite, but it should be "as many as" the natural numbers.

Definition 3.1.8 A set S is **countable** if we can find a surjective function $f : \mathbb{N} \to S$, i.e. we have a way to count *every* element of S using natural numbers.

R It can be seen from the above definition that we do not require f to be *injective*, meaning that we can count S's elements *repetitively*.

Example 3.6 — Countable Sets.

- Finite sets are countable. Why? Consider any finite set $S = \{s_1, \dots, s_n\}$ for some $n \in \mathbb{N}$. Simply take $f : \mathbb{N} \to S$ such that $f(x) = s_x$ if $1 \leq x \leq n$, and $f(x) = s_1$ otherwise. It is easy to see that f is surjective.
- $\mathbb{N}$ is countable. [Left as an exercise.]
- $\mathbb{Z}$ is countable. Why? Let us define $f : \mathbb{N} \to \mathbb{Z}$ such that $f(x) = -\frac{x-1}{2}$ if x is odd, and $f(x) = \frac{x}{2}$ if x is even. It can be shown that this function is surjective. This actually shows that on the set level, a set that is *bigger* than $\mathbb{N}$ *may still be countable*.
- Given a countable set S, every $A \subset S$ is countable. [By definition, left as an exercise.]

We do have some important results with regards to countability of sets, one of which is that countability holds with $\mathbb{N}$-unions.

Theorem 3.1.3 If $\{S_i\}_{i \in \mathbb{N}}$ are a collection of countable sets, $S = \bigcup_{i=1}^{\infty} S_i$ is also countable.

Proof. This proof is a very clever trick to count the elements in the union. We can observe that each element $a \in S$ can be characterized uniquely by (i, x) where i represents which S_i that a is in, and x represents the natural number we give $a \in S_i$ since S_i is countable. Given this $a \in S$ and (i, x) representing a, we consider $m = i + x$, and we define the function $f : \mathbb{N} \to S$ such that

$$f(\frac{(m-1)(m-2)}{2} + i) = a$$

for all $a \in S, m \in \mathbb{N}$. Basically we are counting elements in S in the following order: $s_{11}, s_{12}, s_{21}, s_{13}, s_{22}, s_{31}, \dots$, where s_{11} is the first element we count in S_1 and so on.

Since each $a \in S$ gets a unique (i,x) pair, and each pair corresponds to a unique $\frac{(m-1)(m-2)}{2}+i$ value, we know that f is surjective. Hence, S is countable. $\square$

R In fact, we can try to do the above proof using mathematical induction, but even if we show that $S_n = \bigcup_{i=1}^{n} S_i$ is countable for all $n \in \mathbb{N}$, it is still different (requires more theory) to show that S is countable since S is an infinite union of sets.

Above proof for the theorem also can be used to prove another important theorem, whose proof will left as an exercise:

Theorem 3.1.4 The rational numbers $\mathbb{Q}$ is countable.

Next, what about $\mathbb{R}$? In fact, $\mathbb{R}$ is *not* countable. Let us see why.

Theorem 3.1.5 $\mathbb{R}$ is not countable.

Proof. Suppose by contradiction that $\mathbb{R}$ is countable. Then, we can define a function $f : \mathbb{N} \to \mathbb{R}$ such that f is surjective. Since f is surjective, we can actually define an *injective* inverse function $f^{-1} : \mathbb{R} \to \mathbb{N}$, mapping each number in $\mathbb{R}$ to their counting order, and in fact $f^{-1}(\mathbb{R}) \subset \mathbb{N}$.
Let us consider the numbers $0.2111111\ldots, 0.22111111\ldots, 0.22211111\ldots, 0.222211111\ldots, \cdots$ Each of the numbers are *different* since they has different number of 2's. Let S be the set of these numbers. We know that they all have different counting orders. Also, we can define $g : S \to \mathbb{N}$ such that $g(s) =$"number of 2's in s.", and in fact g is a bijection onto $\mathbb{N}$.
Since we can find a bijective mapping from S to $\mathbb{N}$, it is not possible that $f^{-1}(S) \neq \mathbb{N}$ since we already know f^{-1} is injective on all of $\mathbb{R}$. Why is this? Consider any $h : \mathbb{N} \to \mathbb{N}$ such that $h(n) = f^{-1}(g^{-1}(n))$. Then, if h is not onto, we know that there are $n_1, n_2 \in \mathbb{N}$ such that $h(n_1) = h(n_2)$, however $g^{-1}(n_1) \neq g^{-1}(n_2)$, thus by definition of h, f^{-1} cannot be injective. Therefore, h must be onto $\mathbb{N}$. It is easy to see that $g^*h = f^{-1}|_S$ by definition of h, thus f^{-1} is bijective.
This means $f^{-1}(S) = \mathbb{N}$, then S must be equal to $\mathbb{R}$, by definition of f^{-1}. However, simply consider the number 0.3, which is different from all numbers in S but is still a real number in $\mathbb{R}$. Thus, $S \neq \mathbb{R}$, hence we reached a contradiction. $\mathbb{R}$ is uncountable. $\square$

3.2 Limits

In the last subsection, we understand what is a conversation with $\mathbb{R}$ (functions) and different types of functions, understands the operations on these conversations and how to log them (graphs of functions). In the end, we talked about the concept of countability defined using functions.
In this subsection, we will try do make sense of the idea that "where the conversation is going". In fact, this is the idea similar to what we have for defining sup and inf for $\mathbb{R}$: the idea of "approaching". We will give more concrete definitions of this using *limits*.

Definition 3.2.1 Given a function $f : A \to \mathbb{R}$, $A \subset \mathbb{R}$, and given $a \in \mathbb{R}$ we write $\lim_{x \to a} f(x) = l$ (called as x goes to a, $f(x)$ *approaches* l or the **limit** of f as x goes to a is l) if for all $\varepsilon > 0$, we can find a $\delta > 0$ such that for all $x \in A$, "$0 < |x-a| < \delta \Rightarrow |f(x)-l| < \varepsilon$" is True. We say the limit of f as x goes to a **exists** if we can find such an $l \in \mathbb{R}$, if not, we say that the limit **does not exist (DNE)**.

R The formulation of above definition suggests that depending on what we choose for ε, δ can change dependently. However, for every value of ε, finding at least one δ that works in the definition is required.

It is in fact the most complicated definition so far, because it involves a lot of logical statements and math symbols. I will try to explain the intuition behind the definition of limits.
As we mentioned above, we would like to know "where the conversation was going", in fact "If I am going to say something with emotion a, what would the other person's response emotion look like?". It would be a *successful and accurate* prediction of the other person's response emotion l if I am able to say something close to a that can trigger this person's response to be as close as possible to l. This is exactly the logic behind predicting the emotional flow of the conversation. That's why the limit definition precisely reflects the idea that, if: no matter how close (any $\varepsilon > 0$) I would want, I can say something close to what I am predicting that I would say ($|x-a| < \delta$), and making the other person's response emotion $f(x)$, ε-close to my prediction l. Then: this prediction l must be where this conversation is going (emotion-wise) if I'm going to say a word with emotion a.
With the above intuition, it actually subtly presents the idea that we need to have a sense of "closeness" between two words with some emotion values. In fact, having the idea of closeness of emotions between words is *much more important* than the idea of "what is the emotion of a *single* word", and the idea of "approaching" is very fundamental in calculus and analysis. In many real analysis sources, authors define *closeness (emotional distance)* at the start, and define *norms* based on the idea of closeness. I will do the other way around, and I believe this way is more intuitive with our natural language analogy.

Definition 3.2.2 We can define a map $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that $d(x,y) = |x-y|$ where $|\cdot|$ is a norm defined on $\mathbb{R}$. We call the map d a *metric* on $\mathbb{R}$.

■ **Example 3.7 — Metrics.** ■

-

3.3 Sequences

3.4 Continuity

3.5 Differentiability