Sets and Functions

Sets

A *set* is a collection of objects called *elements*. Curly braces { } are used to display the list of elements explicitly or by description. For example,

$$\{1, 2, 15\}$$

is the set whose elements are the numbers 1, 2, and 15, while

$$\{\text{real numbers greater than 5}\}$$

is the indicated set of real numbers. Elements of a set can be numbers, a vectors, frogs, or anything else.

Two sets are *equal* if they have exactly the same elements. Thus

$$\{\text{cat}, \text{dog}, \text{elephant}\} = \{\text{elephant}, \text{cat}, \text{dog}\}.$$

Sets can have a finite number of elements (as in the examples above) or an infinite number of elements (for example, the set of real numbers).

The empty set. It is also useful to have a notation for a set that contains no elements. This set is called the *empty set*, and is denoted $\emptyset$. One reason this is useful is that it lets one say "Let S be the set of objects with such-and-such property," without knowing ahead of time that there *are* any objects with this property.

Notation for elements.

The symbol "$\in$" is used to indicate that an object belongs to a set; "$\notin$" is used to indicate that an object does not belong. For example,

- The stand-alone notation $x \in S$ is read "x is an element of S" or simply "x is in S".
- The sentence "If $x \in S$, then x is blue" is read "If x is in S, then x is blue."
- The sentence "If $x \notin S$, then x is odd" is read "If x is not in S, then x is odd."
- The sentence "Let $x \in S$" is read "Let x be an element of S".
- The sentence "Let $x \notin S$" is read "Let x not be an element of S".

Notation for subsets.

We say that a set B is a subset of a set A if every element of B also is in A. We write $B \subset A$ in this case (read this as "B is a subset of A" or "B is contained in A").

Examples:

- Every set A has two "trivial" subsets: $A \subset A$ and $\emptyset \subset A$. (Of course these two subsets are the same if A itself is empty!)
- { positive real numbers } $\subset$ { all real numbers }.
- The set $\{1, 2, 3\}$ has exactly the following eight subsets: $\emptyset$, $\{1\}.\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, $\{1, 2, 3\}$. (Notational fine point: if $x \in S$, we don't write $x \subset S$; instead we write $\{x\} \subset S$.)

A common strategy for showing that two sets A and B are equal is to show that each set is a subset of the other. (In fact this is a more precise way to define "equal sets" than the way above. If two sets are infinite, what else would you mean by saying that they have the same elements?)

Set selector notation

When one wants to define a set by some properties of its elements, *set selector* notation is often used. Either a colon or a vertical bar may be used, as below.

$$A = \{\text{doodad} \mid \text{doodad has property } C\}$$

or

$$A = \{\text{doodad} : \text{doodad has property } C\}.$$

In each case, the notation defines a set A. The vertical line and the colon are pronounced "such that". (These symbols *never* mean "equals".) Within the curly brackets, the symbol that appears to the left of the vertical bar (or colon) is the notation for a typical element of A.

Examples.

$\{x : x \text{ is a real number and } 0 < x < 5\}$.

$\{x \mid x \text{ is a real number and } 0 < x < 5\}$.

The colon and vertical bar mean exactly the same thing; it's a matter of personal preference which you use. I use the vertical bar more; the linear algebra textbook by Friedberg, Insel, and Spence uses the colon. Both examples above are read "The set of x such that x is a real number and $0 < x < 5$".

It is common to denote the set of real numbers by $\mathbf{R}$. An alternative way of describing the set above is $\{x \in \mathbf{R} \mid 0 < x < 5\}$. We use the notation $\mathbf{R}^n$ (which is read "R-n", not "R to the n") for the set of ordered n-tuples of real numbers. Thus

$$\begin{aligned}
\mathbf{R}^2 &= \{(x, y) \mid x, y \in \mathbf{R}\}, \\
\mathbf{R}^3 &= \{(x, y, z) \mid x, y, z \in \mathbf{R}\}, \\
\mathbf{R}^n &= \{(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \in \mathbf{R}\},
\end{aligned}$$

etc. It is also common to write these n-tuples vertically instead of horizontally.

More advanced examples of using set selector notation.

Sometimes additional, very broad restrictions on set elements are put to the left of the bar (or colon), as in

$$A = \{\text{persons } P : P \text{ is over 5 feet tall}\}.$$

This tells us that A is a set contained in the "universe" of persons (really just a larger set of which A is a subset), and that what determines whether a given person is in A is whether he/she is over 5 feet tall.

Sometimes additional objects appear in the same sentence as "$A = \ldots$". Objects that are defined or restricted only *outside* the curly braces are *constants* for the given set A. The notation for these constants (e.g. $a_1, \mathbf{v}_2$) is fixed by the notation outside the brackets; you must use the same letters throughout the problem to refer to the same set A. Different values of the constants may distinguish one set A from another, but are constants as far as a single set A is concerned. ("Constant" here does not necessarily mean "number".) It is irrelevant whether these constants appear *before* or *after* the curly braces; all that matters is that they are outside.

Example 1: $A = \{\text{persons } P : P \text{ is over } b \text{ feet tall}\}$, where b is a real number.

Example 2: $A = \{\text{persons } P : P \text{ sometimes wears a } b\}$, where b is an article of clothing.

In each case, the restriction on b occurs outside the brackets, meaning that b is a constant within each set. In example 1, for $b = 5$, we get one specific set A; for $b = 6$, we get another specific set A. Within each of these sets, b is constant. In example 2, when b is a hat we get one specific set A; when b is a tie we get another specific set A. The definition of A thus allows us to consider a lot of sets at the same time, rather than forcing us to treat each possible A separately.

When additional objects in the set-definition sentence occur *within* the curly brackets, to the right of the bar (or colon), the interpretation is different. This is most easily illustrated by examples before giving the general rules.

Example 3: $A = \{\mathbf{v} \in \mathbf{R}^3 \mid \mathbf{v} = c_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \text{ for some } c_1, c_2 \in \mathbf{R}\}$.

In this example, c_1, c_2 are variable scalars that distinguish *one element* of the fixed set A from *another element*; for every choice of the pair (c_1, c_2), we get a (potentially) different element of A.

Example 4: Let $\mathcal{F}(\mathbf{R}, \mathbf{R})$ be the set of functions with domain and target $\mathbf{R}$, and let

$$A = \{f \in \mathcal{F}(\mathbf{R}, \mathbf{R}) : \text{there exist } c_1, c_2 \in \mathbf{R} \text{ such that } f(x) = c_1 \cos x + c_2 \sin x \text{ for all } x \in \mathbf{R}\}.$$

Again, c_1, c_2 are variable scalars that distinguish *one element* of the fixed set A from *another element*; for every choice of the pair (c_1, c_2), we get a (potentially) different element of A. The variable x does not restrict f at all or distinguish one f from another; x is simply a "dummy variable" we temporarily need in order to describe the relation between two functions.

In general, when additional objects in the set-definition sentence occur *within* the curly brackets, to the right of the bar (or colon), and furthermore (i) potentially distinguish one element of A from another, and (ii) are defined or restricted only by words *within* the brackets, then these objects are *variables* within the set A: different values of these variables determine different elements of A. In examples 3 and 4, c_1, c_2 are such objects; in example 4, x is not.

Example 5: $H = \{(x_1, x_2, \ldots, x_n) \in \mathbf{R}^n : a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0\}$, where $a_1, a_2, \ldots, a_n$ are scalars (*i.e.* real numbers) not all zero.

The notation tells us that there are many sets H defined by this notation, one set for each choice of the scalars. (Because no restriction appears on n, we treat n as a constant for a given set H, just as if a definition of n or restriction on n occurs outside the brackets.) So, for example, if $n = 2$ then one possible H is $H_1 = \{(x_1, x_2) \in \mathbf{R}^2 : 5x_1 + 3x_2 = 0\}$; another is $H_2 = \{(x_1, x_2) \in \mathbf{R}^2 : -7x_1 + 6x_2 = 0\}$. The vector $(1, -5/3)$ is in H_1 but not in H_2.

Example 6: Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be arbitrary vectors in a vector space V. Let

$$H = \{\mathbf{v} \in V : \mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n, \text{where } a_1, a_2, \ldots, a_n \text{ are all scalars}\}.$$

In this example, each choice of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ leads to a *different set* H. For a *single* set H, we must treat the $\mathbf{v}_j$ as constants, not introducing any new notation for them. But within H, the scalars a_j are variables, whose choices determine different vectors in the same set H. These variables can be called by any names we want (e.g. $b_1, \ldots, b_n$).

Equations of a geometric object.

If S is a "geometric object"—*e.g.* some subset of $\mathbf{R}^3$, such as a line, plane, or sphere—we sometimes are asked to "find an equation (or equations) for S". This means we are going to describe S by the property that a given point is an element of S if and only if the given equations are satisfied. Sometimes these equations are written in terms of the coordinates of Euclidean space (such as x, y, z); other times they are *parametric*, written in terms of variables that are not themselves coordinates.

Example 1 (non-parametric). *The plane P in $\mathbf{R}^3$ with equation $x+2y+3z = 6$* means

$$P = \{(x, y, z) \in \mathbf{R}^3 \mid x + 2y + 3z = 6\}.$$

Example 2 (parametric). Let $\mathbf{r} = (x, y, z), \mathbf{u} = (-2, 1, 0), \mathbf{v} = (-3, 0, 1)$. *The plane P in $\mathbf{R}^3$ with equation $\mathbf{r} = (6, 0, 0) + t_1\mathbf{u} + t_2\mathbf{v}$* means

$$P = \{\mathbf{r} \in \mathbf{R}^3 \mid \text{there exist } t_1, t_2 \in \mathbf{R} \text{ such that } \mathbf{r} = (6, 0, 0) + t_1\mathbf{u} + t_2\mathbf{v}\}.$$

(In fact, the two planes described in examples 1 and 2 are the same. Why?)

Intersection and union.

The *intersection* of two sets A and B is the set of elements common to both A and B. The intersection is denoted $A \bigcap B$:

$$A \bigcap B = \{x \mid x \in A \textit{ and } x \in B\}.$$

The *union* of two sets A and B is the set of elements contained in A *or* B, (*i.e.* in at least one of the two; in math, "or" is always inclusive). The union is denoted $A \bigcup B$:

$$A \bigcup B = \{x \mid x \in A \textit{ or } x \in B\}.$$

Note the correspondence between the symbols and the words:

$$\bigcap \longleftrightarrow \text{and}$$
$$\bigcup \longleftrightarrow \text{or}$$

Functions

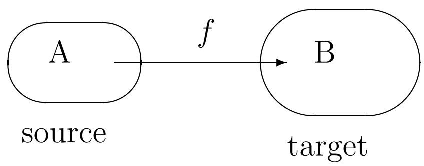
A *function* is an assignment of exactly one element of a specified set, called the *target* or *codomain*, to an element of another set, called the *source* or *domain*. Thus to specify a function, one needs three pieces of information:

- the source A
- the target B
- the rule specifying which element of B gets assigned to each element of A.

A function with source A and target B is said to be a function *from A to B*. The mathematical notation for this is $f : A \to B$, which is read "f from A to B". For f to be a function from A to B, we require that $f(x)$ be defined for every $x \in A$, and that $f(x) \in B$ for every $x \in A$. We do not require that every $y \in B$ be equal to $f(x)$ for some $x \in A$.

It is often useful to picture a function using sets.

Figure 1: A function $f : A \to B$



The term *source* has the same meaning as the word *domain* you used in Calculus 1, but *target* is not the same as *range*. In Calculus 1, the target set of *every* function is $\mathbf{R}$, the set of real numbers, so there's no need to introduce a special word like *target* when this piece of data is the same for all functions. In Calculus 3, when you spoke of vector-valued functions, the target set was usually $\mathbf{R}^3$ or $\mathbf{R}^2$.

In linear algebra, the source and target sets are usually vector spaces, and functions are often called *transformations, maps*, or *mappings*.

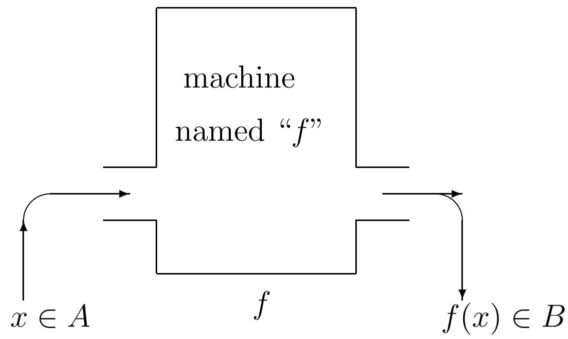
If $f : A \to B$ is a function, we say that f is a *B-valued function on A*. For example, if $B = \mathbf{R}$, we speak of real-valued functions. In Calculus 3, we commonly call $\mathbf{R}^n$-valued functions *vector-valued* functions whether $n = 2, 3$, or anything higher.

The *range* of a function f is set of elements of the target actually "hit" by f, *i.e.* the set of elements of the form $f(x)$ (thus the range is a *subset* of the target). A more pictorial synonym for range is *image*, which is used more often than *range* outside of calculus classes. Thus, if $f : A \to B$ is a function, then

$$\begin{aligned} \text{image}(f) = \text{range}(f) &= \{f(x) \mid x \in A\} \\ &= \{b \in B \mid b = f(x) \text{ for some } x \in A\}. \end{aligned}$$

It is often useful to think of a function as an object (so that for instance one can talk about sets of functions). For this purpose, a useful picture to keep in mind is one in which $f : A \to B$ is viewed as a machine whose input is elements of A and whose output is elements of B:

Figure 2: A function $f : A \to B$



While this picture can be quite useful, it has a drawback: it understates the importance of the sets A and B. Changing either one of these sets changes the function.

But an advantage of this picture is that it allows us to view think of some very complicated functions in more simple terms, avoiding some mental baggage. For example, let S be the set of infinitely differentiable functions with domain $\mathbf{R}$. Define a function $D : S \to S$ by $D(f) = f'$ (*i.e.* the derivative). In your differential equations class, you called D an *operator*. Our more abstract picture shows us that an operator is just another function—it simply has a different source and target from the ones we're more used to.

One-to-one and onto.

A function $f : A \to B$ is called *one-to-one* or *injective* if f takes no two elements of A are taken to the same element of B, *i.e.* if $f(x) \neq f(y)$ whenever $x \neq y$. *One-to-one* is often abbreviated as 1-1.

A function $f : A \to B$ is called *onto* or *surjective* if every element of B is "hit" by f (*i.e.* is in the image of f). Put another way, $f : A \to B$ is onto if image$(f) = B$.

If f is both 1-1 and onto, f is called bijective.

Functions that are injective, surjective, or bijective respectively are called injections, surjections, and bijections.

Difference between a function and a formula.

In your calculus classes, you commonly referred to things like $\sin x, e^x, x^2 + 3x + 2$, and $\sqrt{x}$ as functions. Technically this is wrong: these are *formulas* rather than functions. They are similar to the "machine" part of Figure 2, in which the importance of stating the domain and target is suppressed. This does not mean that everything you learned in calculus is wrong! In your calculus classes, you implicitly used two principles: (i) the target set of every function was $\mathbf{R}$, so never needed to be stated explicitly, and (ii) for the domain of the function, you used the "implied domain" of the formula: the largest set of real numbers for which the formula made sense. For example, in "$\sqrt{x}$", it was always assumed that x was a real number, and the domain was the inteval $[0, \infty) = \{x \in \mathbf{R} \mid x \geq 0\}$. This is called "abuse of notation": notation that is technically incorrect, used when the correct notation would be too cumbersome. For example, in place of "$\sqrt{x}$" one could more correctly (and lengthily) write "the function $f : [0, \infty) \to \mathbf{R}$ defined by $f(x) = \sqrt{x}$".

Abusing notation is not always a bad thing, *provided the writer knows he/she is doing it, and lets the reader know.* One way to avoid abusing notation for commonly used functions is to give these functions a name. For example, "sin" and "cos" are functions; *by definition* their domains and targets are $\mathbf{R}$. (But $\sin(x), \cos(x)$ are technically not functions—they are merely the *output* of the machine in Figure 2.) Another example is the exponential function $\exp : \mathbf{R} \to \mathbf{R}$, which is defined by $\exp(x) = e^x$.

The importance of distinguishing between functions and formulas becomes clearer when one thinks about *inverse functions.* If $f : A \to B$ is a function, we say that a function $g : B \to A$ is *the inverse of* f if both $g(f(x)) = x$ for all $x \in A$ and $f(g(b)) = b$ for all $b \in B$. As you probably learned in Calculus 2, f has an inverse if and only if f is bijective (1-1 and onto). Therefore *the sine function has no inverse*! The function arcsin or $\sin^{-1}$ is *not* the inverse of sine. Instead, it is the inverse of the function

$$\begin{aligned} \text{Sine} : [-\pi/2, \pi/2] &\to [-1, 1], \text{given by} \\ \text{Sine}(x) &= \sin x. \end{aligned}$$

(I've used a capital S to distinguish Sine from sine.) The functions Sine and sine have the same formula, but are different functions: both the domain and target are of Sine are different from those of sine. Similarly, the function sqrt: $[0, \infty) \to \mathbf{R}$ given by $\text{sqrt}(x) = \sqrt{x}$ has no inverse because it is not onto. However, the function Sqrt:$[0, \infty) \to [0, \infty)$ defined using the same formula does have an inverse.