Categories for the uninitiated

Vincent Moreau, IRIF, Université Paris Cité

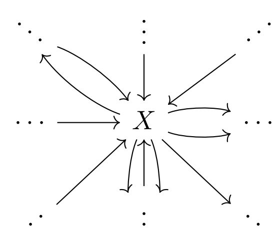
15$^{\text{th}}$ of September 2022, 4pm

Abstract

These are notes for a 1-hour introductory session on category theory accessible to everyone. We introduce categories as generalizations of ordered sets and then focus on the notion of universal property and show through examples how such properties can be used to describe behaviours in different areas of mathematics in a unified way. No prerequisites.

> [I]f mathematics is the science of analogy, the study of patterns, then category theory is the study of patterns of mathematical thought—the "mathematics of mathematics," as Eugenia Cheng [...] has put it.
>
> ---
>
> Infinity Category Theory Offers a Bird's-Eye View of Mathematics, Emily Riehl

Category theory is all about studying objects not by looking at what they individually are, but how they relate to other objects of the same kind. In a way, it is a behaviorist approach to mathematical objects in which the only way to understand an object X is to determine what kind of special place X occupies: each object is immersed in a web or net of relations with all the others.



For instance, let us suppose that we want to understand a given person. One could try to dissect the person, or in a less extreme case to measure his or her body using diverse instruments, make the person do physical and mental tests, etc. In a way, this is a "medical" approach. Another strategy would be to know his friends, family and colleagues and understand what are their relationships to the person, to study his or her social status in the society, his or her interests given all the groups the person may be a member of. This is a more sociological "approach". Categories are all about this second approach.

We first show how this approach is used in the case of posets, and then introduce categories and show how the constructions seen in the first, simpler case are in fact categorical constructions.

1 The case of posets

We first consider posets, namely sets together with a partial order (i.e. a reflexive transitive and anti-symetric relation). For the sake of simplicity, let us take $\mathbb{N}$ with the divisibility relation defined as

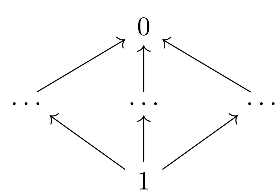
$$\forall m, n \in \mathbb{N}, \qquad m \mid n \iff \exists k \in \mathbb{N},\ n = k \times m.$$

This binary relation $\mid$ links all natural numbers in a certain way, it adds a net of relations between them. It is also a source of information regarding natural numbers as all of them do not behave in the same way regarding divisibility: certain numbers are inferior to others (e.g. $2 \mid 6$) while others cannot be compared in any way (e.g. 3 and 5).

1.1 Minimums and maximums

Some numbers have a special position with respect to divisibility. For instance there is a number which is inferior to all numbers: it is the number 1, as 1 divides all natural numbers (for any n, $n = n \times 1$). In the same way, there is a number which is superior to all numbers: it is 0 (for any n, $0 = 0 \times n$).

We therefore have a net of relations between natural numbers, which is of the form

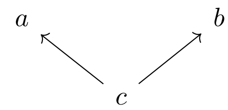


where an arrow from m to n represents the fact that m divides n.

1.2 LCDs and GCDs

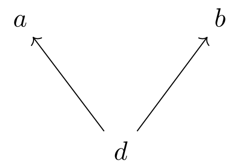
Now, given two natural numbers a and b, we can construct their greatest common divisor $a \wedge b$. This number is the unique natural number verifying the two following properties: (i) it is a natural number c which divides both a and b and (ii) any natural number d which divides both a and b also divides c.

We can represent this situation graphically. With our notaion with arrows, the property (i) becomes

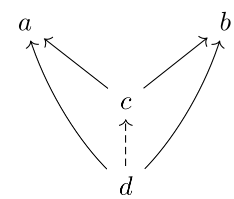


while the property (ii) becomes, for any natural number d,

if



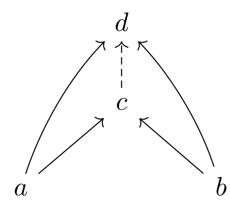
then



where we represent with dots the arrow that the property gives us.

All of this shows that, given two natural numbers a and b, there is a natural number $a \wedge b$ at a special position relative to a and b with respect to divisibility. The definition we have given of the greatest common divisor does not rely on a computation or a formula, but solely on an axiomatized divisibility behavior.

The whole story is identical for the least common multiple if we reverse all the arrows. It can be summarized by the following diagram, where c is $a \vee b$ and d is any natural number:



1.3 The floor and ceiling functions

Consider the two sets $\mathbb{Z}$ and $\mathbb{R}$ with their usual order (not the division anymore). We finish this ordered introduction with the following question: how to represent a real number by an integer in the most faithful way? An answer could be the floor function which associates to any real number $x \in \mathbb{R}$ the greatest integer $\lfloor x \rfloor \in \mathbb{Z}$ which is inferior to x. This integer $\lfloor x \rfloor$ is the unique $n \in \mathbb{Z}$ which verifies the following property:

$$\forall m \in \mathbb{Z}, \qquad m \leq x \quad \iff \quad m \leq n \ .$$

In a way, we are trying to coerce $x \in \mathbb{R}$ into an integer $\lfloor x \rfloor \in \mathbb{Z}$, and we define the latter by asking that the net of inequalities in $\mathbb{Z}$ looks the same as the one in $\mathbb{R}$ when we only consider integers. The integer $\lfloor x \rfloor$ summarizes x from the point of view of integers $\mathbb{Z} \subseteq \mathbb{R}$.

To make this more precise, one has to make the distinction between the two orders, $\leq_{\mathbb{Z}}$ on $\mathbb{Z}$ and $\leq_{\mathbb{R}}$ on $\mathbb{R}$. The two posets $(\mathbb{Z}, \leq_{\mathbb{Z}})$ and $(\mathbb{R}, \leq_{\mathbb{R}})$ are connected by the inclusion map $i : \mathbb{Z} \to \mathbb{R}$ which sends every integer to the same number but seen as real, hence forgetting the fact that it is an integer. With this formalism, one can thus define, for any $x \in \mathbb{R}$, the number $\lfloor x \rfloor \in \mathbb{Z}$ as the unique integer verifying the following property:

$$\forall m \in \mathbb{Z}, \qquad i(m) \leq_{\mathbb{R}} x \quad \Longleftrightarrow \quad m \leq_{\mathbb{Z}} \lfloor x \rfloor \ .$$

The situation may be summarized graphically in the following way:

$$\begin{array}{lccccccccccc}
\text{In } \mathbb{R}\text{:} & \cdots & \longrightarrow & i(m') & \longrightarrow & i(m) & \longrightarrow & x & \longrightarrow & x' & \longrightarrow & \cdots \\
 & & & \uparrow & & \uparrow & \Updownarrow & \downarrow & & \downarrow & & \\
\text{In } \mathbb{Z}\text{:} & \cdots & \longrightarrow & m' & \longrightarrow & m & \longrightarrow & \lfloor x \rfloor & \longrightarrow & \lfloor x' \rfloor & \longrightarrow & \cdots
\end{array}$$

Of course, the ceiling function is defined in a dual way, with the property that $\lceil x \rceil \in \mathbb{Z}$ is the (provably unique) integer such that

$$\forall m \in \mathbb{Z}, \qquad \lceil x \rceil \leq_{\mathbb{Z}} m \quad \Longleftrightarrow \quad x \leq_{\mathbb{R}} i(m) \ .$$

Notice that in the case of the $\lfloor - \rfloor$ function, the i is on the left-hand side of the inequality $\leq_{\mathbb{R}}$ and $\lfloor - \rfloor$ is on the right-hand side of the inequality $\leq_{\mathbb{Z}}$. In the case of $\lceil - \rceil$, it is the contrary: i is on the right-hand side of $\leq_{\mathbb{R}}$ while $\lceil - \rceil$ is on the left-hand side of $\leq_{\mathbb{Z}}$.

All the definitions we have considered here – maximum and minimum elements, least upper and greatest lower bounds of two elements, floor and ceiling functions – can be expressed relatively to any poset, although there might not always exist.

2 Introducing categories

2.1 Informal interlude – Why categories?

We have seen that posets already permit some kind of behaviorist approach to their elements. However, the poset case has some clear limitations. Let us consider, for instance, the poset $\mathbb{N}$ with its usual order.

$$0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \longrightarrow \cdots$$

This poset $\mathbb{N}$ is very abstract, and strangely far from the common intuition about natural numbers. Intuitively, a natural number may be represented by a finite set of "things", whatever this means. The set is represented graphically, and the number it represents is invariant under geometrical transformations.

This representation, for example, gives us a quick and easy intuition about multiple identites of arithmetic. For example, multiplication of two sets representing a and b may be defined as a rectangle with width a and height b.

5 flower petals

10 stones

Therefore, the commutativity of multiplication amounts to a simple rotation of the rectangle, which is a geometrical transformation assumed not to change the number the set represents. The distributivity of multiplication over addition simply amounts to other operations like that.

What about the order? A finite sets of "things" can be considered to be included in another one if there is an injective map from the former to the latter. Again, this is a fairly intuitive idea, which corresponds to the inequality of natural numbers: if a is represented by A and b by B, then a is less than b if and only if A can be injected into B. However, it is clear in that situation that we loose information when we simply state that $a \leq b$: we loose the injection $A \to B$. There may be multiple ones, for example

$$\begin{array}{ccc} \{*_1\} & \longrightarrow & \{*_1, *_2\} \\ *_1 & \longmapsto & *_1 \end{array} \quad \text{and} \quad \begin{array}{ccc} \{*_1\} & \longrightarrow & \{*_1, *_2\} \\ *_1 & \longmapsto & *_2 \end{array}$$

are two different ways to inject the set $\{*_1\}$, representing 1, into the set $\{*_1, *_2\}$, representing 2.

This informal digression on the natural numbers has one goal: to illustrate the fact that, instead of merely saying that two objects are in relation, it is a finer statement to say that there is a special link, the injection in our example, that puts two objects in relation. We consider that a relation

$$a \leq b$$

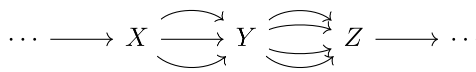
which may be true or false, does not give us as detailed information as we would like. Therefore, we consider instead a set

$$\mathbb{N}(a, b)$$

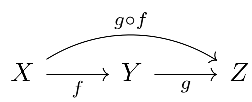
containing all the ways to put a and b in relation (in that order, from a to b). If the set is empty, then a and b are not in relation: otherwise, the elements of the set are the different ways to to relate a with b. This is linked with the idea of proof-relevance: we are not merely interested in the truth value of a statement, but instead in the set of all its potential proofs.

2.2 Formal definition & examples

A category $\mathbf{C}$ consists of objects, here written X, Y, Z, and, for any two objects X and Y, a set $\mathbf{C}(X, Y)$ of elements called the arrows from X to Y. In the rest of the text, we write $f : X \to Y$ to mean that $f \in \mathbf{C}(X, Y)$ when the category $\mathbf{C}$ can be disambiguated from context. Graphically:



Together with objects and sets of arrows comes a composition law $\circ$ whith associates to any arrows $f : X \to Y$ and $g : Y \to Z$ an arrow $g \circ f : X \to Z$.



The composition is required to be associative, i.e. for objects X, Y, Z, W and arrows $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$, there is an equality between the two possible composites:

$$(h \circ g) \circ f = h \circ (g \circ f).$$

A category is also required to have neutral elements for each object X, namely an arrow $1_X : X \to X$ such that, for any two objects X and Y and arrow $f : X \to Y$, we have

$$1_Y \circ f = f = f \circ 1_X.$$

All of this data, namely the objects X, Y, Z, ..., all the sets $\mathbf{C}(X, Y)$ together with the composition law $\circ$ which is associative and has neutral elements, is what is called a category.

If we have two objects X and Y of a category $\mathbf{C}$ and an arrow $f : X \to Y$, then we say that f is an isomorphism if there exists an arrow $g : Y \to X$ such that $g \circ f = 1_X$ and $f \circ g = 1_Y$. We say that two objects X and Y are isomorphic if there exists an isomorphism from X to Y. We write:

$$X \cong Y.$$

It is a general fact that categorical constructions yield objects unique "up to isomorphism", which means that there can be different objects satisfying a given property, which will however be isomorphic. From the categorical point of view, two isomorphic objects X and Y are identified since one can use the isomorphism from X to Y to show that they relate exactly in the same way to other objects.

With our definition, it is easy to see that any poset $(P, \leq)$ induces a category $\mathbf{C}_P$ whose objects are elements of P together with, for any p and q in P, the set

$$\mathbf{C}_P(p,q) = \begin{cases} \{*\} & \text{if } p \leq q \\ \varnothing & \text{otherwise} \end{cases}$$

Therefore, one has a morphism $f : p \to q$ if and only if $p \leq q$, in which case the morphism f is equal to the unique element $*$ of $\mathbf{C}_P(p,q)$. The composition is given by the transitivity of $\leq$; indeed, if $f : p \to q$ (namely $p \leq q$) and $g : q \to r$ (namely $q \leq r$), then $p \leq r$ so the set $\mathbf{C}_P(p,q)$ is non-empty and we let $g \circ f : p \to r$ be its unique element $*$. It is associative, and has unit arrows $1_p : p \to p$ because $p \leq p$ as the relation $\leq$ is reflexive.

We introduce the category **Set** whose objects are sets and whose arrows $f : X \to Y$ are usual set-theoretic functions from X to Y. The composition of arrows is the usual composition of functions, which is associative and has as unit 1_X the identity function Id_X for all objects X.

Last but not least, we introduce the category **Grph** whose objects are directed simple graphs[1], namely sets V together with a binary relation $E \subseteq V \times V$ representing the edges of the graphs. In this category **Grph**, if (V, E) and (V', E') are two graphs, an arrow $f : (V, E) \to (V', E')$ is a set-theoretic function $f : V \to V'$ such that for all v_1 and v_2 in V, we have

$$\text{if} \quad (v_1, v_2) \in E \quad \text{then} \quad (f(v_1), f(v_2)) \in E'$$

which means that f is defined on vertices and sends edges of the input on edges of the output. One can easily show that the composition is well-defined, associative and that the identity functions are the units.

3 Universal properties in our examples

We now reexplore the constructions in posets seen in the first section, but with categorical lenses. We focus first on initial and terminal objects, which are categorical generalizations of minimum and maximum elements of a poset, and then on products and coproducts, which are categorical generalizations of the greatest common divisors and least common multiples. The situation is summarized in the joint table.

Remember that the table does not merely show a connection between posets and categories. As seen in section 2 with the $\mathbf{C}_P$ construction, any poset can be seen as a specific category, so it really is a generalization.

[1]Categoricians usually call *graphs* structures which can have multiple edges between two vertices, also known as multigraphs. They speak of *simple graph* in the case where there is at most one edge from a vertex to another. We prefer to call graph the latter structure, redefined in the text, for the sake of simplicity.

Posets	Categories
minimum	initial object
maximum	terminal object
gcd	product
lcd	coproduct
floor	right adjoint
ceiling	left adjoint

3.1 Initial and terminal objects

Let $\mathbf{C}$ be any category. We say that an object X is initial if, for any object Y of the category, there exists a unique arrow $i_X : X \to Y$, which we represent in the diagram

$$\text{For any object } Y, \qquad X \overset{i_Y}{\dashrightarrow} Y$$

We recall that $\dashrightarrow$ is a shortcut to mean "there exists a unique arrow". In the same way, we say that an object X is terminal if, for any object Y, there exists a unique arrow $t_Y : Y \to X$, which we represent as

$$\text{For any object } Y, \qquad Y \overset{t_Y}{\dashrightarrow} X$$

Notice that the unique arrow starts from X in the initial case, and arrives to X in the terminal case. In general, initial and terminal objects differ.

We talked about the unicity of categorical constructions up to isomorphism. It is now time to prove that in the case of initial objects (the terminal case is the same, reversed). Suppose that you have two initial objects X and X'. By composition of the arrow $i_{X'}$ (given by initiality of X) and i'_X (given by initiality of X'), we obtain an arrow $i'_X \circ i_{X'}$ from X to X. There is another arrow from X to X, namely 1_X. The situation is the following:

$$X \xrightarrow{i_{X'}} X' \xrightarrow{i'_X} X$$

$$X \xrightarrow[1_X]{\qquad\qquad} X$$

However, by initiality of X applied with Y taken to be X itself, there exists a unique arrow $X \to X$, namely $i_X : X \to X$. Therefore, the two abovementionned arrows must be equal to i_X and we obtain that

$$i'_X \circ i_{X'} = 1_X.$$

In the same way, using initiality of X' applied to Y taken to be X' itself, one finds the other equality

$$i_{X'} \circ i'_X = 1_{X'}.$$

We have shown that $i_{X'} : X \to X'$ and $i'_X : X' \to X$ are inverses of each other. Therefore, X and X' are isomorphic so the initial object, if it exists, is unique up to isomorphism.

One can easily verify that, in the category **Set**, the empty set $\varnothing$ is initial and any singleton $\{*\}$ is terminal. Notice that all singletons are isomorphic in the categorical sense as any two singletons are linked by a bijection.

Let us now study the case of **Grph**. Just like in the case of **Set**, one can easily check that the empty graph is initial in **Grph**. However, the terminal object is slightly less clear. It is quite clear that if a graph has no vertex or more than two, it won't be a terminal object as there will be no arrow or multiple ones arriving to it. We try the two graphs with one vertex.

The first graph with one vertex that we consider is the one which has no edge, which we note O_0: formally, O_0 is $(\{*\}, \varnothing)$. For any graph G, it is clear that there is at most one arrow $G \to O_0$ as all such arrows must send the vertices of G on $*$. However, the unicity fails. Indeed, consider for example the case where G is the graph representing an interval, namely G is $(\{v_1, v_2\}, \{(v_1, v_2)\})$. Then, the edge (v_1, v_2) must be sent on an edge of O_0, but there is none, so there cannot be an arrow $G \to O_0$.

We see that the terminal object must have one vertex but must also have a reflexive loop on its unique vertex: we write O_1 this graph, which formally is $(\{*\}, \{(*, *)\})$. Then, for any graph $G = (V, E)$, we consider the unique function $t_G : V \to \{*\}$, which is constant equal to $*$. We now prove that it is an arrow of **Grph**: if (v_1, v_2) is in E, then $(t_G(v_1), t_G(v_2))$ is $(*, *)$, which is an edge of O_1. Therefore, O_1 is terminal in **Grph**.

3.2 Product and coproduct

We now describe what is the cartesian product in **Set** and **Grph**, without being as detailed as for the case of initial and terminal objects.

In **Set**: if X and Y are two sets, their product is the usual product

$$X \times Y = \{(x, y) : x \in X, y \in Y\}$$

and the projections are $\pi_1 : (x, y) \mapsto x$ and $\pi_2 : (x, y) \mapsto y$.

Their coproduct is the disjoint union

$$X \sqcup Y = \{(1, x) : x \in X\} \cup \{(2, y) : y \in Y\}$$

and the inclusions are $\iota_1 : x \mapsto (1, x)$ and $\iota_2 : y \mapsto (2, y)$.

In **Grph**: If $G = (V, E)$ and $G' = (V', E')$ are two graphs, their product is

$$G \times G' = (V \times V', \{((v_1, v'_1), (v_2, v'_2)) : (v_1, v_2) \in E, (v'_1, v'_2) \in E'\})$$

and the projections are the set-theoretic ones which are arrows of **Grph**.

Their coproduct is the disjoint union of vertices and edges

$$G \sqcup G' = (V \sqcup V', E \sqcup E')$$

and the injections are the set-theoretic ones which are arrows of **Grph**.

Notice that, precisely because there can be more arrows of type $H \to G \times G$ than of type $H \to G$ (twice as much, in general), this makes the product $G \times G$ different from G as it occupies a different place in the net of relations:

$$G \times G \ \ncong \ G \ .$$

This is a big difference with the case of partially ordered sets, where we do not count the relations and merely know if one element is less than another. In that setting, the greatest common divisor of a number a and itself is still a:

$$a \vee a \ = \ a \ .$$

It is the "proof-relevance", which we already talked about when introducing categories, which makes this phenomenon possible.

3.3 To make a graph from a set

We have already seen in the first section how to construct an integer from a real. Given a function $i : \mathbb{Z} \to \mathbb{R}$ which forgets the integral nature of an integer, one could define two functions $\lfloor - \rfloor : \mathbb{R} \to \mathbb{Z}$ and $\lceil - \rceil : \mathbb{R} \to \mathbb{Z}$ which were some kind of right and left approximate inverses of i.

One can also be interested in the following question: how to turn a set, an object of **Set**, into a graph, an object of **Grph**. This is an analogous question to the one with numbers. Indeed, we have some forgetful map[2] U : **Grph** $\to$ **Set** which sends a graph $G = (V, E)$ to its underlying set of vertices $UG = V$.

Now, recall that the defining property for the $\lfloor - \rfloor$ function is

$$\forall m \in \mathbb{Z}, \qquad i(m) \leq_{\mathbb{R}} x \qquad \Longleftrightarrow \qquad m \leq_{\mathbb{Z}} \lfloor x \rfloor \ .$$

If we translate this into the language of categories, namely if we write the set of arrows $P(a, b)$ instead of the relation $a \leq_P b$ and replace the equivalence $\Longleftrightarrow$ by the relation $\cong$ of being in bijection, we get

$$\forall m \in \mathbb{Z}, \qquad \mathbb{R}(i(m), x) \quad \cong \quad \mathbb{Z}(m, \lfloor x \rfloor) \ .$$

Now, we want to define a map R : **Set** $\to$ **Grph** which would have the same properties relatively to U : **Grph** $\to$ **Set** than $\lfloor - \rfloor : \mathbb{R} \to \mathbb{Z}$ had relatively to

[2]The assignment U is a functor from the category **Grph** to the category **Set**. This means that it respects the structure of categories, i.e. the composition and neutral elements. We won't use that fact here and will only refer to the fact that U sends a graph on its underlying set of vertices.

$i : \mathbb{Z} \to \mathbb{R}$. For any object X of **Set**, we let RX be the object of **Grph** such that

$$\forall\ G \text{ object of } \mathbf{Grph}, \qquad \mathbf{Set}(UG, X) \quad \cong \quad \mathbf{Grph}(G, RX)\ .$$

We now show that there is indeed such a graph RX. Let us consider $RX = (X, X \times X)$, the complete graph with underlying set X. We verify that for any graph $G = (V, E)$, the two sets $\mathbf{Set}(UG, X)$ and $\mathbf{Grph}(G, RX)$ are in bijection. Let us pose the following function:

$$\Phi \quad : \quad \begin{array}{ccc} \mathbf{Set}(UG, X) & \longrightarrow & \mathbf{Grph}(G, RX) \\ f & \longmapsto & f \end{array}\ .$$

The simplicity of this definition hides some complexity: we must show that the function $f : V \to X$ is an arrow of **Grph** from G to RX. To verify that, we take v_1 and v_2 in V such that $(v_1, v_2) \in E$. Then, $(f(v_1), f(v_2)) \in X \times X$, so f sends adjacent vertices of G on adjacent vertices of RX, so f is an arrow.

It is clear that the map Φ is injective, as an arrow of **Grph** is determined by its action on vertices. It is also surjective as any arrow from G to RX has an underlying map from $UG = V$ to X. Therefore, Φ is a bijection.

We have found a property for defining R from U by analogy with $\lfloor - \rfloor$ and i. The same thing can be done by analogy with $\lceil - \rceil$ and i

$$\forall m \in \mathbb{Z}, \qquad \lceil x \rceil \leq_{\mathbb{Z}} m \quad \iff \quad x \leq_{\mathbb{R}} i(m)\ .$$

in which case we consider an object LX of **Grph** verifying the property

$$\forall\ G \text{ object of } \mathbf{Grph}, \qquad \mathbf{Grph}(LX, G) \quad \cong \quad \mathbf{Set}(X, UG)\ .$$

It turns out that one can check that $LX = (X, \varnothing)$ verifies this property.

To sum up, we have in the case of $i : \mathbb{Z} \to \mathbb{R}$ two functions $\lceil - \rceil, \lfloor - \rfloor : \mathbb{R} \to \mathbb{Z}$ which are some kind of approximate inverses on the left and right of i. In the case of $U : \mathbf{Grph} \to \mathbf{Set}$, we have two assignments $L, R : \mathbf{Set} \to \mathbf{Grph}$ which are the graphs that look the most like X through the lens of U, which are respectively the discrete and the complete graphs.

This concept, although with a bit more technical details, is known as *adjunction* and is a fundamental concept of category theory.

Have fun!