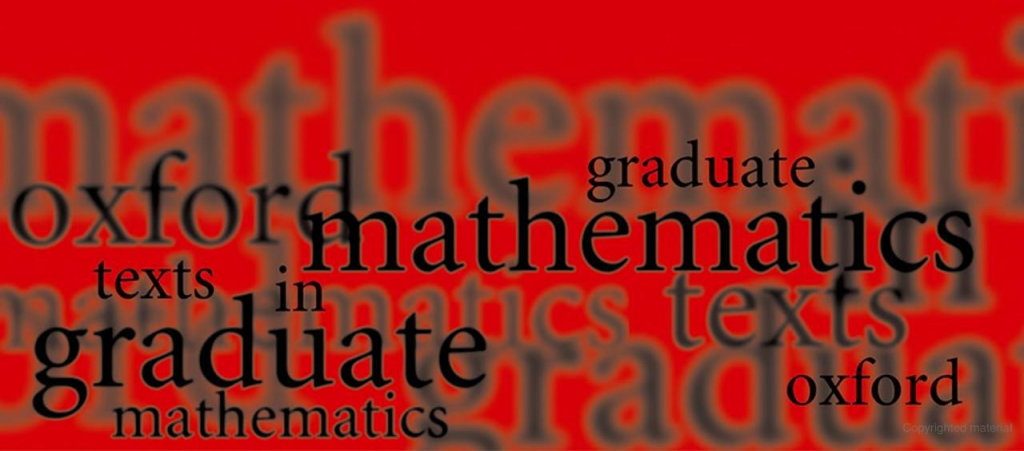
OXFORD
MATHEMATICS

Categories for Quantum Theory: An Introduction

Chris Heunen and Jamie Vicary

OXFORD GRADUATE TEXTS IN MATHEMATICS | 28



Categories for Quantum Theory

An Introduction

CHRIS HEUNEN AND JAMIE VICARY

University of Edinburgh and University of Birmingham

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2019

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2019941499

ISBN 978–0–19–873962–3 (hbk.)
ISBN 978–0–19–873961–6 (pbk.)

DOI: 10.1093/oso/9780198739623.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Contents

0 Basics

Traditional first courses in category theory and quantum computing would prepare the reader with solid foundations for this book. However, not much of that material is truly essential to get the most out of this book. This chapter gives a very brief introduction to category theory, linear algebra and quantum computing, enough to get you going with this book if you have not taken a course in any of these areas before, or perhaps to remind you of some details if you have forgotten them. Everything in this chapter can be found in more detail in many other standard texts (see the Notes at the end of the chapter for references). You could skip this chapter for now, and refer back to it whenever some background is missing.

The material is divided into three sections. Section 0.1 gives an introduction to category theory, and in particular the categories **Set** of sets and functions, and **Rel** of sets and relations. Section 0.2 introduces the mathematical formalism of Hilbert spaces that underlies quantum mechanics, and defines the categories **Vect** of vector spaces and linear maps, and **Hilb** of Hilbert spaces and bounded linear maps. Section 0.3 recalls the basics of quantum theory, including the standard interpretation of states, dynamics and measurement and the quantum teleportation procedure.

0.1 Category Theory

This section gives a brief introduction to category theory. We focus in particular on the category **Set** of sets and functions, and the category **Rel** of sets and relations, and present a matrix calculus for relations. We introduce the idea of commuting diagrams, and define isomorphisms, groupoids, skeletal categories, opposite categories and product categories. We then define functors, equivalences and natural transformations, and also products, equalizers and idempotents.

Categories for Quantum Theory: An Introduction. Chris Heunen and Jamie Vicary, Oxford University Press (2019).
 DOI: 10.1093/oso/9780198739623.001.0001

0.1.1 Categories

Categories are formed from two basic structures: *objects* $A, B, C, \ldots$, and *morphisms* $A \xrightarrow{f} B$ going between objects. In this book, we will often think of an object as a *system*, and a morphism $A \xrightarrow{f} B$ as a *process* under which the system A becomes the system B. Categories can be constructed from almost any reasonable notion of system and process. Here are a few examples:

- physical systems, and physical processes governing them;
- data types in computer science, and algorithms manipulating them;
- algebraic or geometric structures in mathematics, and structure-preserving functions;
- logical propositions, and implications between them.

Category theory is quite different from other areas of mathematics. While a category is itself just an algebraic structure—much like a group, ring or field—we can use categories to organize and understand other mathematical objects. This happens in a surprising way: by neglecting all information about the structure of the objects, and focusing entirely on relationships *between* the objects. Category theory is the study of the patterns formed by these relationships. While at first this may seem limiting, it is in fact empowering, as it becomes a general language for the description of many diverse structures.

Here is the definition of a category.

Definition 0.1. A *category* $\mathbf{C}$ consists of the following data:

- a collection $\mathrm{Ob}(\mathbf{C})$ of *objects*;
- for every pair of objects A and B, a collection $\mathbf{C}(A,B)$ of *morphisms*, with $f \in \mathbf{C}(A,B)$ written $A \xrightarrow{f} B$;
- for every pair of morphisms $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ with common intermediate object, a *composite* $A \xrightarrow{g \circ f} C$;
- for every object A an *identity morphism* $A \xrightarrow{\mathrm{id}_A} A$.

These must satisfy the following properties, for all objects A, B, C, D, and all morphisms $A \xrightarrow{f} B$, $B \xrightarrow{g} C$, $C \xrightarrow{h} D$:

- *associativity*:

$$h \circ (g \circ f) = (h \circ g) \circ f; \tag{0.1}$$

- *identity*:

$$\mathrm{id}_B \circ f = f = f \circ \mathrm{id}_A. \tag{0.2}$$

We will also sometimes use the notation $f : A \longrightarrow B$ for a morphism $f \in \mathbf{C}(A,B)$.

From this definition we see quite clearly that the morphisms are 'more important' than the objects; after all, every object A is canonically represented by its identity

morphism id_A. This seems like a simple point, but it is a significant departure from much of classical mathematics, in which particular structures (like groups) play a more important role than the structure-preserving maps between them (like group homomorphisms.)

Our definition of a category refers to collections of objects and morphisms, rather than sets, because sets are too small in general. The category **Set** defined in Section 0.1.2 illustrates this well, since Russell's paradox prevents the collection of all sets from being a set. However, such size issues will not play a role in this book, and we will use set theory naively throughout. (See the Notes and Further Reading at the end of this chapter for more sophisticated references on category theory.)

0.1.2 The Category **Set**

The most basic relationships between sets are given by functions.

Definition 0.2. For sets A and B, a *function* $A \xrightarrow{f} B$ comprises, for each $a \in A$, a choice of element $f(a) \in B$. We write $f : a \mapsto f(a)$ to denote this choice.

Writing $\emptyset$ for the empty set, the data for a function $\emptyset \longrightarrow A$ can be provided trivially; there is nothing for the 'for each' part of the definition to do. So there is exactly one function of this type for every set A. However, functions of type $A \longrightarrow \emptyset$ cannot be constructed unless $A = \emptyset$. In general there are $|B|^{|A|}$ functions of type $A \longrightarrow B$, where $|-|$ indicates the cardinality of a set.

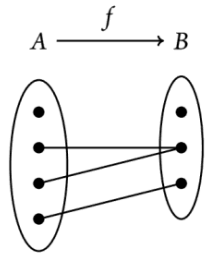
We can now use this to define the category of sets and functions.

Definition 0.3 (Set, FSet). In the category **Set** of sets and functions:

- **objects** are sets $A, B, C, \ldots$;
- **morphisms** are functions $f, g, h, \ldots$;
- **composition** of $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ is the function $g \circ f : a \mapsto g(f(a))$; this is the reason the standard notation $g \circ f$ is not in the other order, even though that would be more natural in some equations such as (0.5);
- **the identity morphism** on A is the function $\mathrm{id}_A : a \mapsto a$.

Write **FSet** for the restriction of **Set** to finite sets.

We will often think of a function $A \xrightarrow{f} B$ in a dynamical way, as indicating how elements of A can evolve into elements of B. This suggests the following sort of picture:



(0.3)

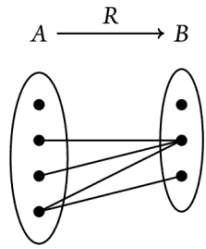
0.1.3 The Category **Rel**

Relations give a more general notion of process between sets.

Definition 0.4. Given sets A and B, a *relation* $A \xrightarrow{R} B$ is a subset $R \subseteq A \times B$.

If elements $a \in A$ and $b \in B$ satisfy $(a, b) \in R$, we often indicate this by writing $a\,R\,b$, or even $a \sim b$ when R is clear. Since a subset can be defined by giving its elements, we can define our relations by listing the related elements, in the form $a_1\,R\,b_1$, $a_2\,R\,b_2$, $a_3\,R\,b_3$ and so on.

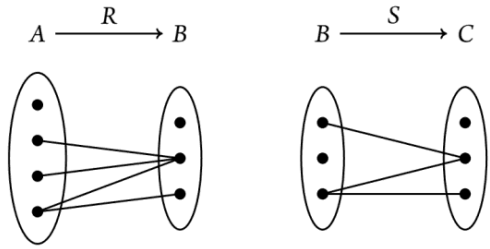
We can think of a relation $A \xrightarrow{R} B$ in a dynamical way, generalizing (0.3):



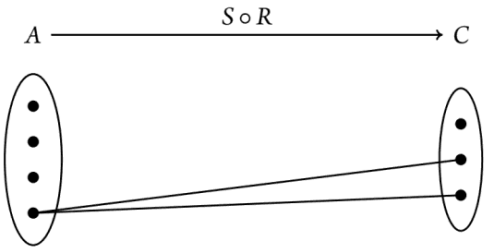
(0.4)

The difference with functions is that this picture indicates interpreting a relation as a kind of nondeterministic classical process: each element of A can evolve into any element of B to which it is related. Nondeterminism enters here because an element of A can relate to more than one element of B, so under this interpretation, we cannot predict perfectly how the system will evolve. An element of A could also be related to *no* elements of B: we interpret this to mean that, for these elements of A, the dynamical process halts. Because of this interpretation, the category of relations is important in the study of nondeterministic classical computing.

Suppose we have a pair of relations, with the codomain of the first equal to the domain of the second:



Our interpretation of relations as dynamical processes then suggests a natural notion of composition: an element $a \in A$ is related to $c \in C$ if there is some $b \in B$ with $a\,R\,b$ and $b\,S\,c$. For the example here, this gives rise to the following composite relation:



This definition of relational composition has the following algebraic form:

$$S \circ R = \{(a, c) \mid \exists b \in B\colon\ aRb \text{ and } bSc\} \subseteq A \times C \tag{0.5}$$

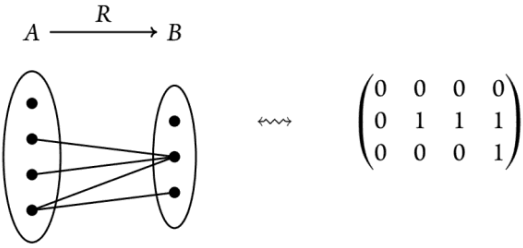
We can write this differently as

$$a\,(S \circ R)\,c \;\Leftrightarrow\; \bigvee_b (bSc \wedge aRb), \tag{0.6}$$

where $\vee$ represents *logical disjunction* (`or`), and $\wedge$ represents *logical conjunction* (`and`). Comparing this with the definition of matrix multiplication, we see a strong similarity:

$$(g \circ f)_{ik} = \sum_j g_{ij} f_{jk} \tag{0.7}$$

This suggests another way to interpret a relation: as a matrix of truth values. For the example relation (0.4), this gives the following matrix, where we write 0 for false and 1 for true:



$$\leftrightsquigarrow \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{0.8}$$

Composition of relations is then just given by ordinary matrix multiplication, with logical disjunction and conjunction replacing + and $\times$, respectively (so that $1 + 1 = 1$).

There is an interesting analogy between quantum dynamics and the theory of relations. First, a relation $A \xrightarrow{R} B$ tells us, for each $a \in A$ and $b \in B$, whether it is *possible* for a to produce b, whereas a complex-valued matrix $H \xrightarrow{f} K$ gives us the *amplitude* for a to evolve to b. Second, relational composition tells us the *possibility* of evolving via an intermediate

point through a sum-of-paths formula, whereas matrix composition tells us the *amplitude* for this to happen.

The intuition we have developed leads to the following category.

Definition 0.5 (Rel, FRel). In the category **Rel** of sets and relations:

- **objects** are sets $A, B, C, \ldots$;
- **morphisms** are relations $R \subseteq A \times B$;
- **composition** of $A \xrightarrow{R} B$ and $B \xrightarrow{S} C$ is the relation

$$\{(a,c) \in A \times C \mid \exists b \in B\colon\ (a,b) \in R, (b,c) \in S\};$$

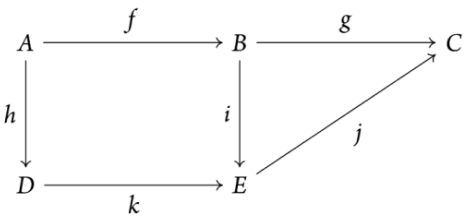
- **the identity morphism** on A is the relation $\{(a,a) \in A \times A \mid a \in A\}$.

Write **FRel** for the restriction of **Rel** to finite sets.

While **Set** is a setting for classical physics, and **Hilb** (to be introduced in Section 0.2) is a setting for quantum physics, **Rel** is somewhere in the middle. It seems like it should be a lot like **Set**, but in fact, its properties are much more like those of **Hilb**. This makes it an excellent test-bed for investigating different aspects of quantum mechanics from a categorical perspective.

0.1.4 Morphisms

It often helps to draw diagrams of morphisms, indicating how they compose. Here is an example:



(0.9)

We say a diagram *commutes* when every possible path from one object in it to another is the same. In the example, this means $i \circ f = k \circ h$ and $g = j \circ i$. It then follows that $g \circ f = j \circ k \circ h$, where we do not need to write parentheses thanks to the associativity equation (0.1). Thus, we have two ways to speak about equality of composite morphisms: by algebraic equations, or by commuting diagrams.

The following terms are very useful when discussing morphisms. The term 'operator' that follows comes from physics.

Definition 0.6 (Domain, codomain, endomorphism, operator). For a morphism $A \xrightarrow{f} B$, its *domain* is the object A, and its *codomain* is the object B. If $A = B$ then we call f an *endomorphism* or *operator*. We sometimes write $\mathrm{dom}(f) = A$ and $\mathrm{cod}(f) = B$.

Definition 0.7 (Isomorphism, retraction). A morphism $A \xrightarrow{f} B$ is an *isomorphism* when it has an *inverse* morphism $B \xrightarrow{f^{-1}} A$ satisfying:

$$f^{-1} \circ f = \mathrm{id}_A \qquad\qquad f \circ f^{-1} = \mathrm{id}_B \tag{0.10}$$

We then say that A and B are *isomorphic*, and write $A \simeq B$. If only the left or right equation of (0.10) holds, then f is called *left-* or *right-invertible*, respectively. A right-invertible morphism is also called a *retraction*.

Lemma 0.8. *If a morphism has an inverse, then this inverse is unique.*

Proof. If g and g' are inverses for f, then:

$$g \overset{(0.2)}{=} g \circ \mathrm{id} \overset{(0.10)}{=} g \circ (f \circ g') \overset{(0.1)}{=} (g \circ f) \circ g' \overset{(0.10)}{=} \mathrm{id} \circ g' \overset{(0.2)}{=} g' \qquad \square$$

Example 0.9. Let us see what isomorphisms are like in our example categories:

- in **Set**, the isomorphisms are exactly the bijections of sets;
- in **Rel**, the isomorphisms are the graphs of bijections: a relation $A \xrightarrow{R} B$ is an isomorphism when there is some bijection $A \xrightarrow{f} B$ such that $aRb \Leftrightarrow f(a) = b$.

The notion of isomorphism leads to some important types of category.

Definition 0.10 (Skeletal category). A category is *skeletal* when any two isomorphic objects are equal.

We will see in Section 0.1.6 that every category is *equivalent* to a skeletal category, which means they encode essentially the same algebraic data.

Definition 0.11 (Groupoid, group). A *groupoid* is a category in which every morphism is an isomorphism. A *group* is a groupoid with one object.

Of course, this definition of group agrees with the ordinary one.

Many constructions with and properties of categories can be easily described in terms of morphisms.

Definition 0.12 (Opposite category). Given a category $\mathbf{C}$, its *opposite* $\mathbf{C}^{\mathrm{op}}$ is a category with the same objects, but with $\mathbf{C}^{\mathrm{op}}(A,B)$ given by $\mathbf{C}(B,A)$. That is, the morphisms $A \longrightarrow B$ in $\mathbf{C}^{\mathrm{op}}$ are morphisms $B \longrightarrow A$ in $\mathbf{C}$.

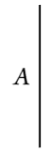
Definition 0.13 (Product category). For categories $\mathbf{C}$ and $\mathbf{D}$, their *product* is a category $\mathbf{C} \times \mathbf{D}$, whose objects are pairs (A,B) of objects $A \in \mathrm{Ob}(\mathbf{C})$ and $B \in \mathrm{Ob}(\mathbf{D})$, and whose morphisms are pairs $(A,B) \xrightarrow{(f,g)} (C,D)$ with $A \xrightarrow{f} C$ and $B \xrightarrow{g} D$.

Definition 0.14 (Discrete category). A category is *discrete* when all the morphisms are identities.

Definition 0.15 (Indiscrete category). A category is *indiscrete* when there is a unique morphism $A \longrightarrow B$ for each two objects A and B.

0.1.5 Graphical Notation

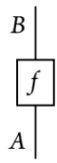
There is a graphical notation for morphisms and their composites. Draw an object A as follows:



(0.11)

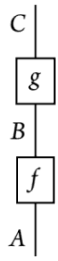
It is just a line. In fact, you should think of it as a picture of the identity morphism $A \xrightarrow{\mathrm{id}_A} A$. Remember, in category theory, the morphisms are more important than the objects.

A morphism $A \xrightarrow{f} B$ is drawn as a box with one 'input' at the bottom, and one 'output' at the top:



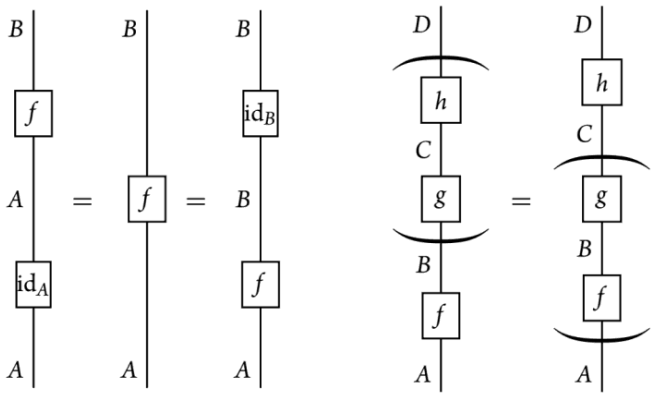
(0.12)

Composition of $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ is then drawn by connecting the output of the first box to the input of the second box:



(0.13)

The identity law $f \circ \mathrm{id}_A = f = \mathrm{id}_B \circ f$ and the associativity law $(h \circ g) \circ f = h \circ (g \circ f)$ then look like:



(0.14)

To make these laws immediately obvious, we choose to not depict the identity morphisms id_A at all and not indicate the bracketing of composites.

The graphical calculus is useful because it 'absorbs' the axioms of a category, making them a consequence of the notation. This is because the axioms of a category are about stringing things together in sequence. At a fundamental level, this connects to the geometry of the line, which is also *one-dimensional*. Of course, this graphical representation is quite familiar: you usually draw it horizontally and call it algebra.

0.1.6 Functors

Remember the motto that in category theory, morphisms are more important than objects. Category theory takes its own medicine here: there is an interesting notion of 'morphism between categories', as given by the following definition.

Definition 0.16 (Functor, covariance, contravariance). Given categories $\mathbf{C}$ and $\mathbf{D}$, a *functor* $F\colon \mathbf{C} \longrightarrow \mathbf{D}$ is defined by the following data:

- for each object $A \in \mathrm{Ob}(\mathbf{C})$, an object $F(A) \in \mathrm{Ob}(\mathbf{D})$;
- for each morphism $A \xrightarrow{f} B$ in $\mathbf{C}$, a morphism $F(A) \xrightarrow{F(f)} F(B)$ in $\mathbf{D}$.

This data must satisfy the following properties:

- $F(g \circ f) = F(g) \circ F(f)$ for all morphisms $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ in $\mathbf{C}$;
- $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$ for every object A in $\mathbf{C}$.

Functors are implicitly *covariant*. There are also *contravariant* versions reversing the direction of morphisms: $F(g \circ f) = F(f) \circ F(g)$. We will only use this covariant definition, and model the contravariant version $\mathbf{C} \longrightarrow \mathbf{D}$ as a covariant functor $\mathbf{C}^{\mathrm{op}} \longrightarrow \mathbf{D}$. A functor between groups is also called a *group homomorphism*; of course this coincides with the usual notion.

We can use functors to give a notion of equivalence for categories.

Definition 0.17 (Equivalence). A functor $F\colon \mathbf{C} \longrightarrow \mathbf{D}$ is an *equivalence* when it is:

- *full*, meaning that the functions $\mathbf{C}(A,B) \longrightarrow \mathbf{D}\big(F(A),F(B)\big)$ given by $f \mapsto F(f)$ are surjective for all $A, B \in \mathrm{Ob}(\mathbf{C})$;
- *faithful*, meaning that the functions $\mathbf{C}(A,B) \longrightarrow \mathbf{D}\big(F(A),F(B)\big)$ given by $f \mapsto F(f)$ are injective for all $A, B \in \mathrm{Ob}(\mathbf{C})$;
- *essentially surjective on objects*, meaning that for each object $B \in \mathrm{Ob}(\mathbf{D})$ there is an object $A \in \mathrm{Ob}(\mathbf{C})$ such that $B \simeq F(A)$.

If two categories are equivalent, then one is just as good as the other for the purposes of doing category theory, even though they might be defined in quite a different way. Nonetheless, one might be much easier to work with than the other, and that is one reason why the notion of equivalence is so useful.

A category $\mathbf{C}$ is a *subcategory* of a category $\mathbf{D}$ when every object of $\mathbf{C}$ is an object of $\mathbf{D}$, every morphism of $\mathbf{C}$ is a morphism of $\mathbf{D}$, and composition and identities in $\mathbf{C}$ are the same as in $\mathbf{D}$. In other words, the inclusion $\mathbf{C} \longrightarrow \mathbf{D}$ is a faithful functor.

Every category has a skeleton, a smaller category with the same algebraic structure, that is equivalent to it.

Definition 0.18 (Skeleton). A *skeleton* of a category $\mathbf{C}$ is a subcategory $\mathbf{S}$ such that every object in $\mathbf{C}$ is isomorphic (in $\mathbf{C}$) to exactly one object in $\mathbf{S}$.

Intuitively, a skeleton is built by restricting the category $\mathbf{C}$ to contain just one object from each isomorphism class. The definition says, in other words, that the inclusion functor $\mathbf{S} \longrightarrow \mathbf{C}$ is an equivalence and that $\mathbf{S}$ is skeletal.

0.1.7 Natural Transformations

Just as a functor is a map between categories, so there is a notion of a map between functors, called a *natural transformation*.

Definition 0.19 (Natural transformation, natural isomorphism). Given functors $F\colon \mathbf{C} \longrightarrow \mathbf{D}$ and $G\colon \mathbf{C} \longrightarrow \mathbf{D}$, a *natural transformation* $\zeta : F \Longrightarrow G$ is an assignment to every object A in $\mathbf{C}$ of a morphism $F(A) \xrightarrow{\zeta_A} G(A)$ in $\mathbf{D}$, such that the following diagram commutes for every morphism $A \xrightarrow{f} B$ in $\mathbf{C}$:

$$\begin{array}{ccc} F(A) & \xrightarrow{\zeta_A} & G(A) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(B) & \xrightarrow[\zeta_B]{} & G(B) \end{array} \tag{0.15}$$

If every component ζ_A is an isomorphism then ζ is called a *natural isomorphism*, and F and G are called *naturally isomorphic*.

Many important concepts in mathematics can be defined in a simple way using functors and natural transformations, such as the following.

Example 0.20. A *group representation* is a functor $\mathbf{G} \longrightarrow \mathbf{Vect}$, where $\mathbf{G}$ is a group regarded as a category with one object (see Definition 0.11.) An *intertwiner* is a natural transformation between such functors.

The notion of natural isomorphism leads to another characterization of equivalence of categories.

Definition 0.21 (Equivalence by natural isomorphism). A functor $F\colon \mathbf{C} \longrightarrow \mathbf{D}$ is an *equivalence* if and only if there exists a functor $G\colon \mathbf{D} \longrightarrow \mathbf{C}$ and natural isomorphisms $G \circ F \simeq \mathrm{id}_{\mathbf{C}}$ and $\mathrm{id}_D \simeq F \circ G$.

A functor is an equivalence by Definition 0.21 just when it is an equivalence by Definition 0.17, and so we abuse terminology mildly, using the word 'equivalence' for both concepts. It is interesting to consider the difference between these definitions: while Definition 0.17 is written in terms of the internal structure of the categories involved, in the form of their objects and morphisms, Definition 0.21 is written in terms of their external context, given by the functors and natural transformations between them. This is a common dichotomy in category theory, with 'internal' concepts often being more elementary and direct, while the associated 'external' perspective, although making use of more sophisticated notions, is often more powerful and elegant. We revisit this external notion of equivalence in Chapter 8, from the perspective of higher category theory.

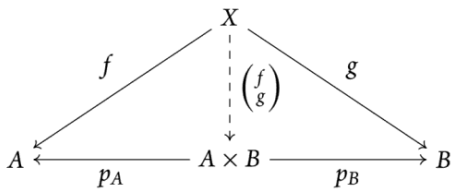
0.1.8 Limits

Limits are recipes for finding objects and morphisms with *universal properties*, with great practical use in category theory. We won't describe the general case here, but just the important special cases of products, equalizers, terminal objects and their dual notions.

To get the idea, it is useful to think about the disjoint union $S + T$ of sets S and T. It is not just a bare set; it comes equipped with functions $S \xrightarrow{i_S} S + T$ and $T \xrightarrow{i_T} S + T$ that show how the individual sets embed into the disjoint union. And furthermore, these functions have a special property: a function $S + T \xrightarrow{f} U$ corresponds exactly to a pair of functions of types $S \xrightarrow{f_S} U$ and $T \xrightarrow{f_T} U$, such that $f \circ i_S = f_S$ and $f \circ i_T = f_T$. The concepts of limit and colimit generalize this observation.

We now define product and coproduct, and also terminal and initial object.

Definition 0.22 (Product, coproduct). Given objects A and B, a *product* is an object $A \times B$ together with morphisms $A \times B \xrightarrow{p_A} A$ and $A \times B \xrightarrow{p_B} B$, such that any two morphisms $X \xrightarrow{f} A$ and $X \xrightarrow{g} B$ allow a unique morphism $\begin{pmatrix} f \\ g \end{pmatrix} : X \longrightarrow A \times B$ with $p_A \circ \begin{pmatrix} f \\ g \end{pmatrix} = f$ and $p_B \circ \begin{pmatrix} f \\ g \end{pmatrix} = g$. The following diagram summarizes these relationships:



A *coproduct* is the dual notion, that reverses the directions of all the arrows in this diagram. Given objects A and B, a coproduct is an object $A + B$ equipped with morphisms $A \xrightarrow{i_A} A + B$ and $B \xrightarrow{i_B} A + B$, such that for any morphisms $A \xrightarrow{f} X$ and $B \xrightarrow{g} X$, there is a unique morphism $(f\ g) : A + B \longrightarrow X$ such that $(f\ g) \circ i_A = f$ and $(f\ g) \circ i_B = g$.

Definition 0.23 (Terminal object, initial object). An object A is *terminal* if for every object X, there is exactly one morphism $X \longrightarrow A$. It is *initial* if for every object X, there is exactly one morphism $A \longrightarrow X$.

A category may not have any of these structures, but if they exist, they are unique up to isomorphism.

Definition 0.24 (Cartesian category). A category is *Cartesian* when it has a terminal object and products of any pair of objects.

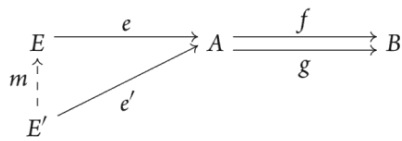
These structures exist in our main example categories.

Example 0.25. Products, coproducts, terminal objects and initial objects take the following forms in our main example categories:

- in **Set**, products are given by the Cartesian product, and coproducts by the disjoint union, any 1-element set is a terminal object, and the empty set is the initial object;
- in **Rel**, products and coproducts are both given by the disjoint union, and the empty set is both the terminal and initial object.

Given a pair of functions $S \xrightarrow{f,g} T$, it is interesting to ask on which elements of S they take the same value. Category theory dictates that we shouldn't ask about elements, but use morphisms to get the same information using a universal property. This leads to the notion of equalizer, a structure that may or may not exist in any particular category.

Definition 0.26. For morphisms $A \xrightarrow{f,g} B$, their *equalizer* is a morphism $E \xrightarrow{e} A$ satisfying $f \circ e = g \circ e$, such that any morphism $E' \xrightarrow{e'} A$ satisfying $f \circ e' = g \circ e'$ allows a unique $E' \xrightarrow{m} E$ with $e' = e \circ m$:



The *coequalizer* of f and g is their equalizer in the opposite category.

Example 0.27. Let us see what equalizers look like in our example categories.

- The categories **Set**, **Vect** and **Hilb** (see Section 0.2) have equalizers for all pairs of parallel morphisms. An equalizer for $A \xrightarrow{f,g} B$ is the set $E = \{a \in A \mid f(a) = g(a)\}$, equipped with its embedding $E \xrightarrow{e} A$; that is, it is the largest subset of A on which f and g agree.
- The category **Rel** does not have all equalizers. For example, consider the relation $R = \{(y,z) \in \mathbb{R}^2 \mid y < z \in \mathbb{R}\} \colon \mathbb{R} \longrightarrow \mathbb{R}$. Suppose $E \colon X \longrightarrow \mathbb{R}$ were an equalizer of R and $\mathrm{id}_{\mathbb{R}}$. Then $R \circ R = \mathrm{id}_{\mathbb{R}} \circ R$, so there is a relation $M \colon \mathbb{R} \longrightarrow X$ with $R = E \circ M$. Now $E \circ (M \circ E) = (E \circ M) \circ E = R \circ E = \mathrm{id}_{\mathbb{R}} \circ E = E$, and since $S = \mathrm{id}_X$ is the unique morphism satisfying $E \circ S = E$, we must have $M \circ E = \mathrm{id}_X$. But then xEy and yMx for some $x \in X$ and $y \in \mathbb{R}$. It follows that $y(E \circ M)y$, that is, $y < y$, which is a contradiction.

A kernel is a special kind of equalizer.

Definition 0.28. A *kernel* of a morphism $A \xrightarrow{f} B$ is an equalizer of f and the *zero* morphism $A \xrightarrow{0} B$ (see Section 2.2.)

A last instance of universal properties is the idea of split idempotents.

Definition 0.29 (Idempotent, splitting). An endomorphism $A \xrightarrow{f} A$ is called *idempotent* when $f \circ f = f$. An idempotent $A \xrightarrow{f} A$ *splits* when there exist an object $\hat{f}$ and morphisms $A \xrightarrow{p_f} \hat{f}$ and $\hat{f} \xrightarrow{i_f} A$ such that the following hold:

$$i_f \circ p_f = f \tag{0.16}$$

$$p_f \circ i_f = \mathrm{id}_{\hat{f}} \tag{0.17}$$

Given such a split idempotent, the injection $\hat{f} \xrightarrow{i_f} A$ gives an equalizer of f and id_A, and the projection $A \xrightarrow{p_f} \hat{f}$ gives a coequalizer of f and id_A.

0.2 Hilbert Spaces

This section introduces the mathematical formalism that underlies quantum theory: complex vector spaces, inner products and Hilbert spaces. We define the categories **Vect** and **Hilb**, and define basic concepts such as orthonormal bases, linear maps, matrices, dimensions and duals of Hilbert spaces. We then introduce the adjoint of a linear map between Hilbert spaces, and define the terms unitary, isometry, partial isometry and positive. We also define the tensor product of Hilbert spaces and introduce the Kronecker product of matrices.

0.2.1 Vector Spaces

A vector space is a collection of elements that can be added to one another, and scaled.

Definition 0.30 (Vector space). A *vector space* is a set V with a chosen element $0 \in V$, an addition operation $+ \colon V \times V \longrightarrow V$, and a scalar multiplication operation $\cdot \colon \mathbb{C} \times V \longrightarrow V$, satisfying the following properties for all $a, b, c \in V$ and $s, t \in \mathbb{C}$:

- *additive associativity*: $a + (b + c) = (a + b) + c$;
- *additive commutativity*: $a + b = b + a$;
- *additive unit*: $a + 0 = a$;
- *additive inverses*: there exists $-a \in V$ such that $a + (-a) = 0$;
- *additive distributivity*: $s \cdot (a + b) = (s \cdot a) + (s \cdot b)$
- *scalar unit*: $1 \cdot a = a$;
- *scalar distributivity*: $(s + t) \cdot a = (s \cdot a) + (t \cdot a)$;
- *scalar compatibility*: $s \cdot (t \cdot a) = (st) \cdot a$.

The prototypical example of a vector space is $\mathbb{C}^n$, the Cartesian product of n copies of the complex numbers.

Definition 0.31 (Linear map, anti-linear map). A *linear map* is a function $f : V \longrightarrow W$ between vector spaces, with the following properties, for all $a, b \in V$ and $s \in \mathbb{C}$:

$$f(a+b) = f(a) + f(b) \tag{0.18}$$

$$f(s \cdot a) = s \cdot f(a) \tag{0.19}$$

An *anti-linear map* is a function that satisfies (0.18), but instead of (0.19), satisfies

$$f(s \cdot a) = s^* \cdot f(a), \tag{0.20}$$

where the star denotes complex conjugation.

Vector spaces and linear maps form a category.

Definition 0.32 (Vect, FVect). In the category **Vect** of vector spaces and linear maps:

- **objects** are complex vector spaces;
- **morphisms** are linear functions;
- **composition** is composition of functions;
- **identity morphisms** are identity functions.

Write **FVect** for the restriction of **Vect** to those vector spaces that are isomorphic to $\mathbb{C}^n$ for some natural number n; these are also called *finite-dimensional*, see Definition 0.34.

Any morphism $f : V \longrightarrow W$ in **Vect** has a kernel, namely the inclusion of $\ker(f) = \{v \in V \mid f(v) = 0\}$ into V. Hence, kernels in the categorical sense coincide precisely with kernels in the sense of linear algebra.

Definition 0.33. The *direct sum* of vector spaces V and W is the vector space $V \oplus W$, whose elements are pairs (a, b) of elements $a \in V$ and $b \in W$, with entrywise addition and scalar multiplication.

Direct sums are both products and coproducts in the category **Vect**. Similarly, the zero-dimensional space is both terminal and initial in **Vect**.

0.2.2 Bases and Matrices

One of the most important structures a vector space can have is a *basis*. A basis gives rise to the notion of dimension of a vector space, and lets us represent linear maps using matrices.

Definition 0.34 (Basis). For a vector space V, a family of elements $\{e_i\}$ is *linearly independent* when every element $a \in V$ can be expressed as a finite linear combination $a = \sum_i a_i e_i$ with *coefficients* $a_i \in \mathbb{C}$ in at most one way. It is a *basis* if additionally any $a \in V$ can be expressed as such a finite linear combination.

Every vector space admits a basis, and any two bases for the same vector space have the same cardinality.

Definition 0.35 (Dimension, finite-dimensionality). The *dimension* of a vector space V, written $\dim(V)$, is the cardinality of any basis. A vector space is *finite-dimensional* when it has a finite basis.

If vector spaces V and W have bases $\{d_i\}$ and $\{e_j\}$, and we fix some order on the bases, we can represent a linear map $V \xrightarrow{f} W$ as the matrix with $\dim(W)$ rows and $\dim(V)$ columns, whose entry at row i and column j is the coefficient $f(d_j)_i$. Composition of linear maps then corresponds to matrix multiplication (0.7). This directly leads to a category.

Definition 0.36 ($\mathbf{Mat}_{\mathbb{C}}$). In the skeletal category $\mathbf{Mat}_{\mathbb{C}}$:

- **objects** are natural numbers $0, 1, 2, \ldots$;
- **morphisms** $n \longrightarrow m$ are complex matrices with m rows and n columns;
- **composition** is given by matrix multiplication;
- **identities** $n \xrightarrow{\mathrm{id}_n} n$ are given by n-by-n matrices with entries 1 on the main diagonal, and 0 elsewhere.

This theory of matrices is 'just as good' as the theory of finite-dimensional vector spaces, made precise by the category theory developed in Section 0.1.

Proposition 0.37. *There is an equivalence of categories* $\mathbf{Mat}_{\mathbb{C}} \longrightarrow \mathbf{FVect}$ *that sends* n *to* $\mathbb{C}^n$ *and a matrix to its associated linear map.*

Proof. Because every finite-dimensional complex vector space H is isomorphic to $\mathbb{C}^{\dim(H)}$, the functor R is essentially surjective on objects. It is full and faithful since there is an exact correspondence between matrices and linear maps for finite-dimensional vector spaces. □

For square matrices, the trace is an important operation.

Definition 0.38 (Trace). For a square matrix with entries m_{ij}, its *trace* is the sum $\sum_i m_{ii}$ of its diagonal entries.

0.2.3 Hilbert Spaces

Hilbert spaces are structures that are built on vector spaces. The extra structure lets us define angles and distances between vectors, and is used in quantum theory to calculate probabilities of measurement outcomes.

Definition 0.39 (Inner product). An *inner product* on a complex vector space V is a function $\langle - | - \rangle : V \times V \longrightarrow \mathbb{C}$ that is:

- *conjugate-symmetric*: for all $a, b \in V$,

$$\langle a | b \rangle = \langle b | a \rangle^*; \tag{0.21}$$

- *linear* in the second argument: for all $a, b, c \in V$ and $s \in \mathbb{C}$,

$$\langle a | s \cdot b \rangle = s \cdot \langle a | b \rangle, \tag{0.22}$$

$$\langle a | b + c \rangle = \langle a | b \rangle + \langle a | c \rangle; \tag{0.23}$$

- *positive definite*: for all $a \in V$,

$$\langle a | a \rangle \geq 0, \tag{0.24}$$

$$\langle a | a \rangle = 0 \Rightarrow v = 0. \tag{0.25}$$

Definition 0.40 (Norm). For a vector space with inner product, the *norm* of an element v is $\|v\| = \sqrt{\langle v | v \rangle}$, a nonnegative real number.

The complex numbers carry a canonical inner product:

$$\langle s | t \rangle = s^* t \tag{0.26}$$

The induced norm satisfies the triangle inequality $\|a + b\| \leq \|a\| + \|b\|$ by virtue of the Cauchy–Schwarz inequality $|\langle a | b \rangle|^2 \leq \langle a | a \rangle \cdot \langle b | b \rangle$, that holds in any vector space with an inner product. Thanks to these properties, it makes sense to think of $\|a - b\|$ as the distance between vectors a and b.

A Hilbert space is an inner product space in which it makes sense to add infinitely many vectors in certain cases.

Definition 0.41 (Hilbert space). A *Hilbert space* is a vector space H with an inner product that is *complete* in the following sense: if a sequence $v_1, v_2, \ldots$ of vectors satisfies $\sum_{i=1}^{\infty} \|v_i\| < \infty$, then there is a vector v such that $\|v - \sum_{i=1}^{n} v_i\|$ tends to zero as n goes to infinity.

Every finite-dimensional vector space with inner product is necessarily complete. Any vector space with an inner product can be completed to a Hilbert space by formally adding the appropriate limit vectors.

There is a notion of bounded map between Hilbert spaces that makes use of the inner product structure. The idea is that for each map there is some maximum amount by which the norm of a vector can increase.

Definition 0.42 (Bounded linear map). A linear map $f : H \longrightarrow K$ between Hilbert spaces is *bounded* when there exists a number $r \in \mathbb{R}$ such that $\|f(a)\| \leq r \cdot \|a\|$ for all $a \in H$.

Every linear map between finite-dimensional Hilbert spaces is bounded.

Hilbert spaces and bounded linear maps form a category. This category will be the main example throughout the book to model phenomena in quantum theory.

Definition 0.43 (Hilb, FHilb). In the category **Hilb** of Hilbert spaces and bounded linear maps:

- **objects** are Hilbert spaces;
- **morphisms** are bounded linear maps;

- **composition** is composition of linear maps as ordinary functions;
- **identity morphisms** are given by the identity linear maps.

Write **FHilb** for the restriction of **Hilb** to finite-dimensional Hilbert spaces.

This definition is perhaps surprising, especially in finite dimensions: since every linear map between Hilbert spaces is bounded, **FHilb** is an equivalent category to **FVect**. In particular, the inner products play no essential role. We will see in Section 2.3 how to model inner products categorically, using the idea of *daggers*.

Hilbert spaces have a more discerning notion of basis.

Definition 0.44 (Basis, orthogonal basis, orthonormal basis). For a Hilbert space H, an *orthogonal basis* is a family of elements $\{e_i\}$ with the following properties:

- they are *pairwise orthogonal*, that is, $\langle e_i | e_j \rangle = 0$ for all $i \neq j$;
- every element $a \in H$ can be written as an infinite linear combination of e_i; that is, there are *coefficients* $a_i \in \mathbb{C}$ for which $\| a - \sum_{i=1}^{n} a_i e_i \|$ tends to zero as n goes to infinity.

It is *orthonormal* when additionally $\langle e_i | e_i \rangle = 1$ for all i.

Any orthogonal family of elements is linearly independent. For finite-dimensional Hilbert spaces, the ordinary notion of basis as a vector space, as given by Definition 0.34, is still useful. Hence, once we fix (ordered) bases on finite-dimensional Hilbert spaces, linear maps between them correspond to matrices, just as with vector spaces. For infinite-dimensional Hilbert spaces, however, having a basis for the underlying vector space is rarely mathematically useful.

If two vector spaces carry inner products, we can give an inner product to their direct sum, leading to the direct sum of Hilbert spaces.

Definition 0.45 (Direct sum). The *direct sum* of Hilbert spaces H and K is the vector space $H \oplus K$, made into a Hilbert space by the inner product $\langle (a_1, b_1) | (a_2, b_2) \rangle = \langle a_1 | a_2 \rangle + \langle b_1 | b_2 \rangle$.

Direct sums provide both products and coproducts for the category **Hilb**. Hilbert spaces have the good property that any closed subspace can be complemented. That is, if the inclusion $U \hookrightarrow V$ is a morphism of **Hilb** satisfying $\|u\|_U = \|u\|_H$, then there exists another inclusion morphism $U^\perp \hookrightarrow V$ of **Hilb** with $V = U \oplus U^\perp$. Explicitly, $U^\perp$ is the *orthogonal subspace* $\{a \in V \mid \forall b \in U \colon \langle a | b \rangle = 0\}$.

0.2.4 Adjoint Linear Maps

The inner product gives rise to the *adjoint* of a bounded linear map.

Definition 0.46. For a bounded linear map $f \colon H \longrightarrow K$, its *adjoint* $f^\dagger \colon K \longrightarrow H$ is the unique linear map with the following property, for all $a \in H$ and $b \in K$:

$$\langle f(a) | b \rangle = \langle a | f^\dagger(b) \rangle. \tag{0.27}$$

The existence of the adjoint follows from the Riesz representation theorem for Hilbert spaces, which we do not cover here. It follows immediately from (0.27) by uniqueness of adjoints that they also satisfy the following properties:

$$(f^\dagger)^\dagger = f; \tag{0.28}$$
$$(g \circ f)^\dagger = f^\dagger \circ g^\dagger; \tag{0.29}$$
$$\mathrm{id}_H{}^\dagger = \mathrm{id}_H. \tag{0.30}$$

Taking adjoints is an anti-linear operation.

Adjoints give rise to various specialized classes of linear maps.

Definition 0.47. A bounded linear map $H \xrightarrow{f} K$ between Hilbert spaces is:

- *self-adjoint* when $f = f^\dagger$;
- a *projection* when $f = f^\dagger$ and $f \circ f = f$;
- *unitary* when both $f^\dagger \circ f = \mathrm{id}_H$ and $f \circ f^\dagger = \mathrm{id}_K$;
- an *isometry* when $f^\dagger \circ f = \mathrm{id}_H$;
- a *partial isometry* when $f^\dagger \circ f$ is a projection;
- and *positive* when $f = g^\dagger \circ g$ for some bounded linear map $H \xrightarrow{g} K$.

The following notation is standard in the physics literature.

Definition 0.48 (Bra-ket). Given an element $a \in H$ of a Hilbert space, its *ket* $\mathbb{C} \xrightarrow{|a\rangle} H$ is the linear map $s \mapsto sa$. Its *bra* $H \xrightarrow{\langle a|} \mathbb{C}$ is the linear map $b \mapsto \langle a|b\rangle$.

You can check that $|a\rangle^\dagger = \langle a|$:

$$\left(\mathbb{C} \xrightarrow{|a\rangle} H \xrightarrow{\langle b|} \mathbb{C}\right) = \left(\mathbb{C} \xrightarrow{\langle b|\circ|a\rangle} \mathbb{C}\right) = \left(\mathbb{C} \xrightarrow{\langle b|a\rangle} \mathbb{C}\right) \tag{0.31}$$

The final expression identifies the number $\langle b|a\rangle$ with the linear map $1 \mapsto \langle b|a\rangle$. Thus the inner product (or 'bra-ket') $\langle b|a\rangle$ decomposes into a bra $\langle b|$ and a ket $|a\rangle$. Originally due to Paul Dirac, this is traditionally called *Dirac notation*.

The correspondence between $|a\rangle$ and $\langle a|$ leads to the notion of a dual space.

Definition 0.49. For a Hilbert space H, its *dual Hilbert space* H^* is the vector space $\mathbf{Hilb}(H, \mathbb{C})$.

A Hilbert space is isomorphic to its dual in an anti-linear way: the map $H \longrightarrow H^*$ given by $|a\rangle \mapsto \varphi_a = \langle a|$ is an invertible anti-linear function. The inner product on H^* is given by $\langle \varphi_a | \varphi_b \rangle_{H^*} = \langle a|b\rangle_H$, and makes the function $|a\rangle \mapsto \langle a|$ bounded.

Some bounded linear maps support a notion of trace.

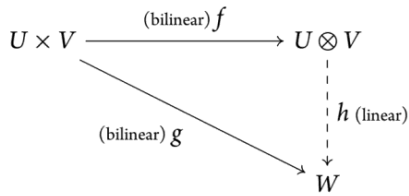
Definition 0.50 (Trace, trace class). When it converges, the *trace* of a positive linear map $f : H \longrightarrow H$ is given by $\mathrm{Tr}(f) = \sum \langle e_i | f(e_i)\rangle$ for any orthonormal basis $\{e_i\}$, in which case the map is called *trace class*.

If the sum converges for one orthonormal basis, then with effort you can prove that it converges for all orthonormal bases, and that the trace is independent of the chosen basis. In the finite-dimensional case, the trace defined in this way agrees with the matrix trace of Definition 0.38.

0.2.5 Tensor Products

The tensor product is a way to make a new vector space out of two given ones. With some work the tensor product can be constructed explicitly, but it is only important for us that it exists, and is defined up to isomorphism by a universal property. If U, V and W are vector spaces, a function $f\colon U \times V \longrightarrow W$ is called *bilinear* when it is linear in each variable; that is, when the function $u \mapsto f(u,v)$ is linear for each $v \in V$, and the function $v \mapsto f(u,v)$ is linear for each $u \in U$.

Definition 0.51. The *tensor product of vector spaces* U and V is a vector space $U \otimes V$ together with a bilinear function $f\colon U \times V \longrightarrow U \otimes V$ such that for every bilinear function $g\colon U \times V \longrightarrow W$ there exists a unique linear function $h\colon U \otimes V \longrightarrow W$ such that $g = h \circ f$.



Note that $U \times V$ is not itself a vector space, so it doesn't make sense to ask if f or g are linear. The function f usually stays anonymous and is written as $(a,b) \mapsto a \otimes b$. It follows that arbitrary elements of $U \otimes V$ take the form $\sum_{i=1}^{n} s_i a_i \otimes b_i$ for $s_i \in \mathbb{C}$, $a_i \in U$ and $b_i \in V$. The tensor product also extends to linear maps. If $f_1\colon U_1 \longrightarrow V_1$ and $f_2\colon U_2 \longrightarrow V_2$ are linear maps, there is a unique linear map $f_1 \otimes f_2\colon U_1 \otimes U_2 \longrightarrow V_1 \otimes V_2$ that satisfies $(f_1 \otimes f_2)(a_1 \otimes a_2) = f_1(a_1) \otimes f_2(a_2)$ for $a_1 \in U_1$ and $a_2 \in U_2$. In this way, the tensor product becomes a functor $\otimes\colon \mathbf{Vect} \times \mathbf{Vect} \longrightarrow \mathbf{Vect}$.

Definition 0.52. The *tensor product of Hilbert spaces* H and K is the Hilbert space $H \otimes K$ built by taking tensor product of the underlying vector spaces, giving it the inner product $\langle a_1 \otimes b_1 \,|\, a_2 \otimes b_2 \rangle = \langle a_1 \,|\, a_2 \rangle_H \cdot \langle b_1 \,|\, b_2 \rangle_K$, then completing it. If $H \xrightarrow{f} H'$ and $K \xrightarrow{g} K'$ are bounded linear maps, then so is the continuous extension of the tensor product of linear maps to a function that we again call $f \otimes g\colon H \otimes K \longrightarrow H' \otimes K'$. This gives a functor $\otimes\colon \mathbf{Hilb} \times \mathbf{Hilb} \longrightarrow \mathbf{Hilb}$.

If $\{e_i\}$ is an orthonormal basis for Hilbert space H, and $\{f_j\}$ is an orthonormal basis for K, then $\{e_i \otimes f_j\}$ is an orthonormal basis for $H \otimes K$. So when H and K are finite-dimensional, there is no difference between their tensor products as vector spaces and as Hilbert spaces.

Definition 0.53 (Kronecker product). When finite-dimensional Hilbert spaces H_1, H_2, K_1, K_2 are equipped with fixed ordered orthonormal bases, linear maps $H_1 \xrightarrow{f} K_1$ and $H_2 \xrightarrow{g} K_2$ can be written as matrices. Their tensor product $H_1 \otimes H_2 \xrightarrow{f \otimes g} K_1 \otimes K_2$ corresponds to the following block matrix, called their *Kronecker product*:

$$(f \otimes g) := \begin{pmatrix} (f_{11}g) & (f_{12}g) & \cdots & (f_{1n}g) \\ (f_{21}g) & (f_{22}g) & \cdots & (f_{2n}g) \\ \vdots & \vdots & \ddots & \vdots \\ (f_{m1}g) & (f_{m2}g) & \cdots & (f_{mn}g) \end{pmatrix}. \tag{0.32}$$

0.3 Quantum Information

Quantum information theory studies the information processing capabilities of quantum systems, using the mathematical abstractions of Hilbert spaces and linear maps.

0.3.1 State Spaces

Classical computer science often considers systems to have a finite set of states. An important simple system is the *bit*, with state space given by the set $\{0, 1\}$. Quantum information theory instead assumes that systems have state spaces given by finite-dimensional Hilbert spaces. The quantum version of the bit is the qubit.

Definition 0.54. A *qubit* is a quantum system with state space $\mathbb{C}^2$.

A *pure state* of a quantum system is given by a vector $v \in H$ in its associated Hilbert space. Such a state is *normalized* when the vector in the Hilbert space has norm 1:

$$\langle a | a \rangle = 1 \tag{0.33}$$

In particular, a complex number of norm 1 is called a phase. A pure state of a qubit is therefore a vector of the form

$$a = \begin{pmatrix} s \\ t \end{pmatrix}$$

with $s, t \in \mathbb{C}$, which is normalized when $|s|^2 + |t|^2 = 1$. In Section 0.3.4 we will encounter a more general notion of state, called a mixed state. However, when our meaning is clear, we'll often just say *state* instead of pure state.

When performing computations in quantum information, we often use the following privileged basis.

Definition 0.55 (Computational basis, Z basis). For the Hilbert space $\mathbb{C}^n$, the *computational basis*, or *Z basis* is the orthonormal basis given by the following vectors:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \qquad \cdots \qquad |n-1\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \tag{0.34}$$

This orthonormal basis is no better than any other, but it is useful to fix a standard choice. Every state $a \in \mathbb{C}^n$ can be written in terms of the computational basis; for a qubit, we can write $a = s|0\rangle + t|1\rangle$ for some $s, t \in \mathbb{C}$. The following alternative qubit basis also plays an important role.

Definition 0.56. The *X basis* for a qubit $\mathbb{C}^2$ is given by the following states:

$$|+\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$
$$|-\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

Processing quantum information takes place by applying unitary maps $H \xrightarrow{f} H$ to the Hilbert space of states. Such a map will take a normalized state $a \in H$ to a normalized state $f(a) \in H$. An example of a unitary map is the *X gate* represented by the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, which acts as $|0\rangle \mapsto |1\rangle$ and $|1\rangle \mapsto |0\rangle$ on the computational basis states of a qubit.

0.3.2 Compound Systems and Entanglement

Given two quantum systems with state spaces given independently by Hilbert spaces H and K, as a joint system their overall state space is $H \otimes K$, the tensor product of the two Hilbert spaces (see Section 0.2.5). This is a postulate of quantum theory. As a result, state spaces of quantum systems grow large very rapidly: a collection of n qubits will have a state space isomorphic to $\mathbb{C}^{2^n}$, requiring 2^n complex numbers to specify its state vector exactly. In contrast, a classical system consisting of n bits can have its state specified by a single binary number of length n.

In quantum theory, (pure) product states and (pure) entangled states are defined as follows.

Definition 0.57 (Product state, entangled state). For a compound system with state space $H \otimes K$, a *product state* is a state of the form $a \otimes b$ with $a \in H$ and $b \in K$. An *entangled state* is a state not of this form.

The definition of product and entangled state also generalizes to systems with more than two components. When using Dirac notation, if $|a\rangle \in H$ and $|b\rangle \in K$ are chosen states, we will often write $|ab\rangle$ for their product state $|a\rangle \otimes |b\rangle$.

The following family of entangled states plays an important role in quantum information theory.

Definition 0.58 (Bell state). The *Bell basis* for a pair of qubits with state space $\mathbb{C}^2 \otimes \mathbb{C}^2$ is the orthonormal basis given by the following states:

$$|\text{Bell}_0\rangle = \tfrac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$
$$|\text{Bell}_1\rangle = \tfrac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$$
$$|\text{Bell}_2\rangle = \tfrac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$$
$$|\text{Bell}_3\rangle = \tfrac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

The state $|\text{Bell}_0\rangle$ is often called 'the Bell state', and is very prominent in quantum information. The Bell states are *maximally entangled*, meaning that they induce an extremely strong correlation between the two systems involved (see Definition 0.72).

0.3.3 Pure States and Measurements

For a quantum system in a pure state, the most basic notion of measurement is a *projection-valued measure*. Quantum theory is a set of rules that says what happens to the quantum state when a projection-valued measurement takes place, and the probabilities of the different outcomes. Recall from Definition 0.47 that projections are maps satisfying $p = p^\dagger = p \circ p$.

Definition 0.59. A finite family of linear maps $H \xrightarrow{f_i} H$ is *complete* when the following holds:

$$\sum_i f_i = \text{id}_H \tag{0.35}$$

Definition 0.60. A family of linear maps $H \xrightarrow{f_i} H$ is *orthogonal* when for any $i \neq j$, the following holds:

$$f_i \circ f_j = 0 \tag{0.36}$$

Definition 0.61 (Projection-valued measure, nondegenerate). A *projection-valued measure (PVM)* on a Hilbert space H is a finite family of projections $H \xrightarrow{p_i} H$ that are complete and orthogonal. A PVM is *nondegenerate* when $\text{Tr}(p_i) = 1$ for all i.

In this definition of PVM, the orthogonality property is actually redundant; that is, a complete family of projections is necessarily also orthogonal. For simplicity, however, we include the orthogonality requirement here directly. Also note that while our PVMs are finite, in general infinite PVMs are possible; for simplicity, we focus on the finite case.

Lemma 0.62. *For a finite-dimensional Hilbert space, nondegenerate projection-valued measures correspond to orthonormal bases, up to phase.*

Proof. For an orthonormal basis $|i\rangle$, define a nondegenerate PVM by $p_i = |i\rangle\langle i|$. Conversely, since projections p have eigenvalues 1, if $\text{Tr}(p) = 1$ then p must have rank one; that is, there is a ket $|i\rangle$ such that $p = |i\rangle\langle i|$, unique up to multiplication by a complex phase $e^{i\theta}$. □

A projection-valued measure, when applied to a Hilbert space, will have a unique outcome, given by one of the projections. This outcome will be probabilistic, with distribution described by the Born rule, defined next. Dirac notation is often extended to self-adjoint bounded linear functions $H \xrightarrow{f} K$ between Hilbert spaces, writing $\langle a|f|b\rangle$ for $\langle a|f(b)\rangle = \langle f(a)|b\rangle$.

Definition 0.63 (Born rule). For a projection-valued measure $\{p_i\}$ on a system in a normalized state $a \in H$, the *probability of outcome* i is $\langle a|p_i|a\rangle$.

The definition of a projection-valued measure guarantees that the total probability across all outcomes is 1:

$$\sum_i \langle a|p_i|a\rangle \overset{(0.18)}{=} \langle a|\left(\sum_i p_i\right)|a\rangle \overset{(0.35)}{=} \langle a|a\rangle \overset{(0.33)}{=} 1 \tag{0.37}$$

After a measurement, the new state of the system is $p_i(a)$, where p_i is the projection corresponding to the outcome that occured. This part of the standard interpretation is called the *projection postulate*. Note that this new state is not necessarily normalized. If the new state is not zero, it can be normalized in a canonical way, giving $p_i(a)/\|p_i(a)\|$.

Given some classical information and some quantum information, it is often the case that we want to apply a unitary operator to the quantum information, in a way that depends on the classical information.

Definition 0.64 (Controlled operation). Given a Hilbert space H and a set S, a *controlled operation* is a choice for all $s \in S$ of a unitary $U_s \colon H \longrightarrow H$.

0.3.4 Mixed States and Measurements

Suppose there is a machine that produces a quantum system with Hilbert space H. The machine has two buttons: one that will produce the system in state $a \in H$ and another that will produce it in state $b \in H$. You receive the system that the machine produces, but you cannot see it operating; all you know is that the operator of the machine flips a fair coin to decide which button to press. Taking into account this uncertainty, the state of the system that you receive cannot be described by an element of H; the system is in a more general type of state, called a *mixed state*.

Definition 0.65 (Density matrix, normalized). A *density matrix* on a Hilbert space H is a positive map $H \xrightarrow{m} H$. A density matrix is *normalized* when $\mathrm{Tr}(m) = 1$. (Warning: a density matrix is not a matrix in the sense of Definition 0.36.)

Recall from Definition 0.47 that m is positive when there exists some g with $m = g^\dagger \circ g$. Density matrices are more general than pure states, since every pure state $a \in H$ gives rise to a density matrix $m = |a\rangle\langle a|$ in a canonical way. This last piece of Dirac notation is the projection onto the line spanned by the vector a.

Definition 0.66 (Pure state, mixed state). A density matrix $m \colon H \longrightarrow H$ is *pure* when $m = |a\rangle\langle a|$ for some $a \in H$; generally, it is *mixed*.


image
not
available

Definition 0.72. A pure state $a \in H \otimes K$ is *maximally entangled* when tracing out either H or K from $|a\rangle\langle a|$ gives a maximally mixed state; explicitly this means the following, for some $s, t \in \mathbb{C}$:

$$\mathrm{Tr}_H(|a\rangle\langle a|) = s \cdot \mathrm{id}_K \qquad \mathrm{Tr}_K(|a\rangle\langle a|) = t \cdot \mathrm{id}_H \tag{0.40}$$

When $|a\rangle$ is normalized, its trace will be a normalized density matrix, so $s = 1/\dim(H)$ and $t = 1/\dim(K)$.

Up to unitary equivalence there is only one maximally entangled state for each system, as the following lemma shows; its proof will follow from Theorem 3.50.

Lemma 0.73. *Any two maximally entangled states $a, b \in H \otimes K$ are related by $(f \otimes \mathrm{id}_K)(a) = b$ for a unique unitary $H \xrightarrow{f} H$.* □

0.3.5 Decoherence

By Lemma 0.62, every nondegenerate projection-valued measure $\{p_1, \ldots, p_n\}$ on a Hilbert space H corresponds (up to a phase) to an orthonormal basis $\{|1\rangle, \ldots, |n\rangle\}$ for H via $p_i = |i\rangle\langle i|$, and hence induces n pure states of H. We may regard this as a *controlled preparation*: depending on some classical data $i = 1, \ldots, n$, we prepare state $|i\rangle$. Consider how this controlled preparation composes with a measurement in the same basis.

If we start with some classical information, use it to prepare a quantum system, and then immediately measure, we should end up with the same classical information we started with. Indeed, according to the Born rule of Definition 0.63, the probability of getting outcome j after preparing state i is:

$$\langle j|p_i|j\rangle = \langle j|i\rangle\langle i|j\rangle = |\langle i|j\rangle|^2, \tag{0.41}$$

which is 1 for $i = j$ but 0 for $i \neq j$.

The other way around is conceptually less straightforward: if you measure a quantum system, yielding a piece of classical data, and then immediately use that to prepare a state of a quantum system, what do you get? Well, supposing that the quantum system starts in a mixed state given by a density matrix $H \xrightarrow{m} H$ with $m = \sum_{ij} c_{ij}|i\rangle\langle j|$, the measurement results in outcome $|i\rangle$ with probability $\mathrm{Tr}(p_i m) = \langle i|m|i\rangle = c_{ii}$, so the state eventually prepared is

$$\sum_i c_{ii}|i\rangle\langle i|. \tag{0.42}$$

The nondiagonal elements of the density matrix m have vanished, and the mixed state has become a convex combination of pure states. This process is called *decoherence*. Any quantum state undergoes decoherence constantly as it interacts with its environment. It takes extremely good experimental control to keep a quantum state from decohering rapidly.