



**University of Minho**  
School of Engineering

Bruna Filipa Martins Salgado

**Metric  $\lambda$ -calculus with conditionals:  
from quantum to probabilistic  
and beyond**





**University of Minho**  
School of Engineering

Bruna Filipa Martins Salgado

**Metric  $\lambda$ -calculus with conditionals:  
from quantum to probabilistic  
and beyond**

Master's Dissertation  
Master in Physics Engineering

Work carried out under the supervision of  
**Renato Jorge Araújo Neves**



# Copyright and Terms of Use for Third Party Work

This dissertation reports on academic work that can be used by third parties as long as the internationally accepted standards and good practices are respected concerning copyright and related rights.

This work can thereafter be used under the terms established in the license below.

Readers needing authorization conditions not provided for in the indicated licensing should contact the author through the RepositóriUM of the University of Minho.

## License granted to users of this work:



**CC BY**

<https://creativecommons.org/licenses/by/4.0/>



# Acknowledgements

Write your acknowledgements here. Do not forget to mention the projects and grants that you have benefited from while doing your research, if any. Ask your supervisor about the specific textual format to use. (Funding agencies are quite strict about this.)





# **Statement of Integrity**

I hereby declare having conducted this academic work with integrity.

I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, Braga, July 2025

Bruna Filipa Martins Salgado



# Abstract

Noisy intermediate-scale quantum (NISQ) computers are expected to operate with severely limited hardware resources. Precisely controlling qubits in these systems comes at a high cost, is susceptible to errors, and faces scarcity challenges. Therefore, error analysis is indispensable for the design, optimization, and assessment of NISQ computing. Nevertheless, the analysis of errors in quantum programs poses a significant challenge. The overarching goal of the M.Sc. project is to provide a fully-fledged quantum programming language on which to study metric program equivalence in various scenarios, such as in quantum algorithmics and quantum information theory.

**Keywords** approximate equivalence,  $\lambda$ -calculus, metric equations



## Resumo

Escrever aqui o resumo (pt)

**Palavras-chave** palavras, chave, aqui, separadas, por, vírgulas



# Contents

<b>Acronyms</b>	<b>xvii</b>
<b>Notation</b>	<b>xix</b>
<b>I Foundations</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Motivation and Context . . . . .	3
1.2 Goals . . . . .	5
1.3 Document Structure . . . . .	5
<b>2 Metric Lambda Calculus</b>	<b>7</b>
2.1 The Lambda Calculus . . . . .	8
2.2 Syntax . . . . .	9
2.2.1 Type system . . . . .	9
2.2.2 (Raw)Terms . . . . .	10
2.2.3 Free and Bound Variables . . . . .	10
2.2.4 Term formation rules . . . . .	11
2.2.5 $\alpha$ -equivalence . . . . .	14
2.2.6 Substitution . . . . .	15
2.2.7 Properties . . . . .	17
2.2.8 Equations-in-context . . . . .	19
2.2.9 Interlude: Booleans . . . . .	21
2.2.10 Metric equational system . . . . .	26
2.3 Category theory . . . . .	28
2.3.1 Categories . . . . .	29

2.3.2	Produts and coproducts . . . . .	33
2.3.3	Functors . . . . .	39
2.3.4	Natural Tranformations . . . . .	41
2.3.5	Equivalence of Categories . . . . .	42
2.3.6	Adjointns . . . . .	43
2.3.7	Monoidal categories . . . . .	45
2.4	Interpretation . . . . .	47
2.4.1	Classical Semantics . . . . .	48
2.4.2	Semantics of metric equations . . . . .	52
<b>3</b>	<b>Metric <math>\lambda</math>-calculus with conditionals</b>	<b>57</b>
3.1	Syntax . . . . .	57
3.2	Interpretation . . . . .	58
3.3	Lattice Theory Preliminaries . . . . .	58
3.4	Soundeness and Completeness . . . . .	59
3.5	A small example: Met . . . . .	62
3.6	Interlude:Booleans - Part 2 . . . . .	63
3.7	Extensionality of the copairing . . . . .	64
3.8	Some remarks on the metric equation chosen . . . . .	64
<b>II</b>	<b>Applications</b>	<b>67</b>
<b>4</b>	<b>Probabilistic Programming</b>	<b>69</b>
4.1	Banach spaces . . . . .	69
4.1.1	Normed spaces . . . . .	69
4.1.2	Banach spaces . . . . .	70
4.2	Category Ban . . . . .	71
4.3	Measure theory . . . . .	71
4.3.1	What is measure theory? . . . . .	71
4.3.2	Measurable spaces and measures . . . . .	73
4.3.3	Spaces of Measures . . . . .	76
4.4	Examples . . . . .	77



<b>5</b>	<b>Quantum computation</b>	<b>79</b>
5.1	Linear Algebra . . . . .	79
5.1.1	Vector spaces . . . . .	79
5.1.2	Linear operators . . . . .	81
5.1.3	Inner product . . . . .	81
5.1.4	Norm and normed spaces . . . . .	83
5.1.5	Eigenvectors and eigenvalues . . . . .	83
5.1.6	Spectral theorem . . . . .	84
5.1.7	Important classes of operators/matrices . . . . .	84
5.1.8	Useful norms . . . . .	86
5.1.9	Tensor Products and Direct Sums of Complex spaces . . . . .	86
5.2	Quantum Computing Preliminaries . . . . .	88
5.2.1	The 2-Dimensional Hilbert Space . . . . .	88
5.2.2	Multi-qubit States . . . . .	90
5.2.3	Unitary operators and Measurements . . . . .	92
5.2.4	Density operators . . . . .	94
5.2.5	Quantum Channels . . . . .	96
5.2.6	Quantum circuits . . . . .	100
5.2.7	No-cloning theorem . . . . .	101
5.3	Functional Analysis . . . . .	102
5.3.1	Hilbert Spaces . . . . .	102
5.3.2	Topology . . . . .	102
5.4	$W^*$ -Algebras . . . . .	105
5.5	Categories for (first-order) quantum computation . . . . .	105
5.6	Examples . . . . .	105
<b>6</b>	<b>Conclusions and future work</b>	<b>107</b>
6.1	Conclusions . . . . .	107
6.2	Prospect for future work . . . . .	107



## List of Figures

1	Term formation rules of affine lambda calculus. . . . .	12
2	Equations-in-context for affine lambda calculus . . . . .	19
3	Metric equational system . . . . .	27
4	Judgment interpretation . . . . .	49
5	Metric equation for conditionals . . . . .	57
6	Bloch sphere representation of a qubit . . . . .	89



# Acronyms

**NISQ** Noisy Intermediate-Scale Quantum [3](#), [4](#), [5](#)

**BNF** Backus-Naur Form [9](#), [10](#)

**CPTP** Completely Positive Trace-Preserving [97](#), [98](#), [101](#)

**OSR** Operator Sum Representation [98](#)



# Notation

$FV(v)$  Set of free variables of a term  $v$ . 10

$v : \mathbb{A}$  Typed term. 11

$\Gamma, \Delta, E$  Typical names for typing contexts. 11

$\Gamma \triangleright v : \mathbb{A}$  Typing judgement. 11

$v[w/x]$  Substitution of a variable  $x$  for a term  $w$  in a term  $v$ . 15

$\Gamma \triangleright v = w : \mathbb{A}$  Equation-in-context. 19

$t =_\epsilon s$  Metric equation. 26

$\|\cdot\|_{\text{op}}$  Operator norm. 69

$\|\cdot\|_\pi$  Projective norm. 71

$V \widehat{\otimes}_\pi W$  Projective tensor product of  $V$  and  $W$ . 71

$\otimes_{\text{meas}}$  Product measure. 75

$\mathcal{M}\mathbb{R}$  Banach space of finite Borel measures on  $\mathbb{R}$ . 76

$\mathbb{R}$  Space of real numbers 79

$\mathbb{C}^n$   $n$ -dimensional complex space. 79

$\mathbb{C}^{n \times m}$  Space of complex matrices of dimension  $n \times m$ . 79

$R, V, W$  Typical names for vector spaces. 80

**dim** ( $V$ ) Dimension of vector space  $V$ . 80

$\langle \cdot, \cdot \rangle$  Inner product. 82

$\overline{(-)}$  Complex conjugate operation. 82

$\|\cdot\|$  Norm of an arbitrary vector. 83

$d(x, y)$  distance between vectors  $x$  and  $y$ . 83

id Identity operator 84

$(-)^{\dagger}$  Adjoint operation. 84

$U$  Typical name for a unitary operator. 85

$\rho$  Typical designation for a density matrix. 85

$\|\cdot\|_2$  Euclidean norm. 86

$\|\cdot\|_1$  Trace norm. 86

$\|\cdot\|_{L^1}$   $L^1$  norm. 86

$\|\cdot\|_{L^\infty}$   $L^\infty$  norm. 86

$(-)^{\otimes n}$  N-fold tensor product. 87

$|\psi\rangle$  Quantum state. Also known as ket. 88

$\langle\psi|$   $|\psi\rangle^{\dagger}$ . Also known as bra. 88

$\langle\psi|\phi\rangle$  Inner product between states  $|\psi\rangle$  and  $|\phi\rangle$ . 88

$|\psi\rangle \otimes |\phi\rangle$  Tensor product of states  $|\psi\rangle$  and  $|\phi\rangle$ . 90

$|\psi\rangle |\phi\rangle$  Tensor product of states  $|\psi\rangle$  and  $|\phi\rangle$ . 90

$|\psi\phi\rangle$  Tensor product of states  $|\psi\rangle$  and  $|\phi\rangle$ . 90

$X$  Pauli operator  $\sigma_x$ . 92

$Z$  Pauli operator  $\sigma_z$ . 92

$H$  Hadamard gate 92

$P$  Phase-shift gate 93

**CNOT** Controlled Not gate 93



$\|\cdot\|_{\diamond}$  Diamond norm. [99](#)

$\mathcal{H}, \mathcal{K}$  Typical names for Hilbert spaces. [102](#)

$\mathcal{H} \otimes_2 \mathcal{K}$  Hilbert space tensor product of  $\mathcal{H}$  and  $\mathcal{K}$ . [102](#)



# **Part I**

## **Foundations**



## Chapter 1

# Introduction

### 1.1 Motivation and Context

Quantum computing dates back to 1982 when Nobel laureate Richard Feynman proposed the idea of constructing computers based on quantum mechanics principles to efficiently simulate quantum phenomena [1].

The field has since evolved into a multidisciplinary research area that combines quantum mechanics, computer science, and information theory. Quantum information theory, in particular, is based on the idea that if there are new physics laws, there should be new ways to process and transmit information. In classical information theory, all systems (computers, communication channels, etc.) are fundamentally equivalent, meaning they adhere to consistent scaling laws. These laws, therefore, govern the ultimate limits of such systems. For instance, if the time required to solve a particular problem, such as the factorization of a large number, increases exponentially with the size of the problem, this scaling behavior remains true irrespective of the computational power available. Such a problem, growing exponentially with the size of the object, is known as a "difficult problem". However, as demonstrated by Peter Shor, the use of a quantum computer with a sufficient number of quantum bits (qubits) could significantly accelerate the factorization of large numbers [2].

This advancement poses a significant threat to the security of confidential data transmitted over the Internet, as the RSA algorithm is based on the computational difficulty of factorizing large numbers.

The quantum computing paradigm holds immense promise, as evidenced by this compelling result in computational complexity theory. While hardware advancements have brought the scientific community closer to realizing this potential, the ultimate goal is yet to be accomplished. A **Noisy Intermediate-Scale Quantum (NISQ)** computer equipped with 50-100

qubits may surpass the capabilities of current classical computers, yet the impact of quantum noise, such as decoherence in entangled states, imposes limitations on the size of quantum circuits that can be executed reliably [3]. Unfortunately, general-purpose error correction techniques [4–6] consume a substantial number of qubits, making it difficult for NISQ devices to make use of them in the near term. For instance, the implementation of a single logical qubit may require between  $10^3$  and  $10^4$  physical qubits [7]. As a result, it is unreasonable to expect that the idealized quantum algorithm will run perfectly on a quantum device, instead only a mere approximation will be observed.

To reconcile quantum computation with NISQ computers, quantum compilers perform transformations for error mitigation [8] and noise-adaptive optimization [9]. Additionally, current quantum computers only support a restricted, albeit universal, set of quantum operations. As a result, nonnative operations must be decomposed into sequences of native operations before execution [10], [11]. In general, perfect computational universality is not sought, but only the ability to approximate any quantum algorithm, with a preference for minimizing the use of additional gates beyond the original requirements. The assessment of these compiler transformations necessitates a comparison of the error bounds between the source and compiled quantum programs. Additionally, in quantum information theory, it is essential to account for errors arising from malicious attacks or noisy channels [12].

This suggests the development of appropriate notions of approximate program equivalence, *in lieu* of the classical program equivalence and underlying theories that typically hinge on the idea that equivalence is binary, *i.e.* two programs are either equivalent or they are not [13].

As previously noted, Shor’s algorithm has played a pivotal role in sparking heightened interest within the scientific community toward quantum computing research. Several quantum programming languages have surfaced over the past 25 years [14, 15]. These include imperative languages such as Qiskit [16] and Silq [17], as well as functional languages such as Quipper [18] and Q# [19]. On one hand, the design of quantum programming languages is strongly oriented towards implementing quantum algorithms. On the other hand, the definition of functional paradigmatic languages or functional calculi serves as a valuable tool for delving into theoretical aspects of quantum computing, particularly exploring the foundational basis of quantum computation [20].

Most of the current research on algorithms and programming languages assumes that ad-

addressing the challenge of noise during program execution will be resolved either by the hardware or through the implementation of fault-tolerant protocols designed independently of any specific application [21]. As previously stated, this assumption is not realistic in the NISQ era. Nonetheless, there have been efforts to address the challenge of approximate program equivalence in the quantum setting.

[22] and [23] reason about the issue of noise in a quantum while-language by developing a deductive system to determine how similar a quantum program is from its idealised, noise-free version. The former introduces the  $(Q, \lambda)$ -diamond norm which analyzes the output error given that the input quantum state satisfies some quantum predicate  $Q$  to degree  $\lambda$ . However, it does not specify any practical method for obtaining non-trivial quantum predicates. In fact, the methods used in [22] cannot produce any post conditions other than  $(I, 0)$  (i.e., the identity matrix  $I$  to degree 0, analogous to a “true” predicate) for large quantum programs. The latter specifically addresses and delves into this aspect.

An alternative approach was explored in [24], using linear  $\lambda$ -calculus as basis – i.e programs are written as linear  $\lambda$ -terms – which has deep connections to both logic and category theory [25], [26]. A notion of approximate equivalence is then integrated in the calculus via the so-called diamond norm, which induces a metric (roughly, a distance function) on the space of quantum programs (seen semantically as completely positive trace-preserving super-operators) [[12]]. The authors argue that their deductive system allows to compute an approximate distance between two quantum programs easily as opposed to computing an exact distance “semantically” which tends to involve quite complex operators. Some positive results were achieved in this setting, but much remains to be done.

Dar mais ênfase ao lambda calculus -> Procurar livros para inspiração

Colocar o que fiz em termos mais leigos

Maybe posso na mesma discorrer um pouco sobre a parte quantica, fazendo o par-entesis que é a área em que tenho mais experiencia

## 1.2 Goals

## 1.3 Document Structure

Nota sobre conhecimentos q assumo

Nota sobre conhecimentos q tenho -> conhecimento superficial o qual apresento, não mais de que isso

Colocar um frase bonitinha sobre como isto acaba por ser tb um resumo do que eu aprendi e portanto são apresentados os conceitos de raiz



## Chapter 2

# Metric Lambda Calculus

Booleanos  $\rightarrow$  Acrescentar reparo do  $\text{dis}(w) = *$

The Lambda Calculus, developed by Church and Curry in the 1930s, serves as a formal language capturing the key attribute of higher-order functional languages, treating functions as first-class citizens, allowing them to be passed as arguments [27]. Moreover, lambda calculus has been proven to be universal in the sense that any computable function can be represented as an expression within the language [28]. Beyond its foundational aspects, this calculus incorporates extensions for modeling side effects, including probabilistic or non-deterministic behaviors and shared memory. Higher-order functions form a pivotal abstraction in practical programming languages such as LISP, Scheme, ML, and Haskell.

This chapter introduces the metric lambda calculus as presented in [24] drawing also from [29–31]. The metric lambda calculus integrates notions of approximation into the equational system of affine lambda calculus, a variant of lambda calculus that restricts each variable to being used at most once. This chapter offers a brief insight into lambda calculus and an overview of the syntax, metric equational system and interpretation of the metric lambda calculus. It also includes an exposition of the definitions and results from category theory used throughout the remainder of the thesis, based on [32–34]. These categorical notions play a central role, as programs are interpreted within distributive symmetric monoidal (closed) categories. It is worth noting that we explicitly prove certain results concerning conditionals; although these results are well known, their proofs appear to be absent from the existing literature. Additionally, we illustrate the usefulness of the “traditional” equational system by utilizing it to demonstrate that the terms  $\text{inl}(*)$  and  $\text{inr}(*)$  possess certain properties that are characteristic of booleans in classical logic. For a more detailed study of lambda calculus theory, the reader is referred to [27].

## 2.1 The Lambda Calculus

The concept of a function takes a central role in the lambda calculus. But what exactly is a function? In most mathematics, the “functions as graphs” paradigm is the most elegant and appropriate framework for understanding functions. Within this paradigm, each function  $f$  has a fixed domain  $X$  and a fixed codomain  $Y$ . The function  $f$  is then a subset of  $X \times Y$  that satisfies the property that for each  $x \in X$  there is a unique  $y \in Y$  such that  $(x, y) \in f$ . Two functions  $f$  and  $g$  are equal if they yield the same output on each input, that is if  $f(x) = g(x)$  for all  $x \in X$ . This perspective is known as the extensional view of functions, as it emphasizes that the only observable property of a function is how it maps inputs to outputs.

On the other hand, the “functions as rules” paradigm is more appropriate within computer science. In this context, defining a function involves specifying a rule or procedure for computing the function. Such a rule is often expressed in the form of a formula, for example,  $f(x) = x^2$ . As with the mathematical paradigm, two functions are considered extensionally equal if they exhibit the same input-output behavior. However, this view also introduces the notion of intensional equality: two functions are intensionally equal if they are defined by (essentially) the same formula.

In the lambda calculus, functions are described explicitly as formulae. The function  $f : x \mapsto f(x)$  is represented as  $\lambda x.f(x)$ . Applying a function to an argument is done by juxtaposing the two expressions. For instance consider the function  $f : x \mapsto x + 1$ , to compute  $f(2)$  one writes  $(\lambda x.x + 1)(2)$ .

The expression of higher-order functions - functions whose inputs and/or outputs are themselves functions- in a simple manner is an essential feature of lambda calculus. For example, the composition operator  $f, g \mapsto f \circ g$  is written as  $\lambda f.\lambda g.\lambda x.f(g(x))$ . Considering the functions  $f : x \mapsto x^2$  and  $g : x \mapsto x + 1$ , to compute  $(f \circ g)(2)$  one writes

$$(\lambda f.\lambda g.\lambda x.f(g(x)))(\lambda x.x^2)(\lambda x.x + 1)(2).$$

As mentioned above, within the “functions as rules” paradigm, is not always necessary to specify the domain and codomain of a function in advance. For instance, the identity function  $f : x \mapsto x$ , can have any set  $X$  as its domain and codomain, provided that the domain and codomain are the same. In this case, one says that  $f$  has type  $X \rightarrow X$ . In the case of the composition operator,  $h = \lambda f.\lambda g.\lambda x.f(g(x))$ , the domain and codomain of the functions  $f$  and  $g$  must match. Specifically,  $f$  can have any set  $X$  as its domain and any set  $Y$

as its codomain, provided that  $Y$  is the domain of  $g$ . Similarly,  $g$  can have any set  $Z$  as its codomain. Thus,  $h$  has type

$$(X \rightarrow Y) \rightarrow (Y \rightarrow Z) \rightarrow (X \rightarrow Z).$$

This flexibility regarding domains and codomains enables operations on functions that are not possible in ordinary mathematics. For instance, if  $f = \lambda x.x$  is the identity function, then one has that  $f(x) = x$  for any  $x$ . In particular, by substituting  $f$  for  $x$ , one obtains  $f(f) = (\lambda x.x)(f) = f$ . Note that the equation  $f(f) = f$  is not valid in conventional mathematics, as it is not permissible, due to set-theoretic constraints, for a function to belong to its own domain.

Nevertheless this remarkable aspect of lambda calculus, this work focuses on a more restricted version of the lambda calculus, known as the simply-typed lambda calculus. Here, each expression is always assigned a type, which is very similar to the situation in mathematics. A function may only be applied to an argument if the argument's type aligns with the function's expected domain. Consequently, terms such as  $f(f)$  are not allowed, even if  $f$  represents the identity function.

## 2.2 Syntax

### 2.2.1 Type system

As previously mentioned, this work focuses on the simply-typed lambda calculus, this work focuses on the simply-typed lambda calculus, where each lambda term is assigned a *type*. Unlike sets, types are *syntactic* objects, meaning they can be discussed independently of their elements. One can conceptualize types as names or labels for set. The definition of the grammar of types for affine lambda calculus is as follows, where  $G$  represents a set of ground types, is given by the following **Backus-Naur Form (BNF)** [35].

$$\mathbb{A} ::= X \in G \mid \mathbb{I} \mid \mathbb{A} \otimes \mathbb{A} \mid \mathbb{A} \oplus \mathbb{A} \mid \mathbb{A} \multimap \mathbb{A} \quad (2.1)$$

Note that this is an inductive definition. Ground types are things such as booleans, integers, and so forth. The type  $\mathbb{I}$  is the unit/empty type with only one element. The type  $\mathbb{A} \otimes \mathbb{A}$  corresponds to the tensor of two types. The type  $\mathbb{A} \oplus \mathbb{A}$  can be seen as the coproduct/sum of types. Finally, the type  $\mathbb{A} \multimap \mathbb{B}$  is the type of linear maps one type to another.

### 2.2.2 (Raw)Terms

The expressions of the lambda calculus are called lambda terms. In the simply-typed lambda calculus, each lambda term is assigned a type. The terms without the specification of a type are called *raw typed lambda terms*. The grammar of *raw typed lambda terms* is given by the **BNF** below.

$$\begin{aligned} v, v_1, \dots, v_n, w \quad ::= \quad & x \mid f(v_1, \dots, v_n) \mid * \mid (\lambda x : \mathbb{A}.v) \mid (vw) \mid v \otimes w \mid \\ & \text{pm } v \text{ to } x \otimes y.w \mid v \text{ to } *.w \mid \text{dis}(v) \mid \text{inl}(v) \mid \text{inr}(v) \mid \\ & \text{case } v \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow w; \text{inr}_{\mathbb{A}}(y) \Rightarrow u \} \end{aligned}$$

Here  $x$  ranges over an infinite set of variables.  $f \in \Sigma$ , where  $\Sigma$  corresponds to a class of sorted operation symbols, and  $f(v_1, \dots, v_n)$  corresponds to the application of the function  $f$  to the arguments  $v_1, \dots, v_n$ . The symbol  $*$  is the unit element of the type  $\mathbb{I}$ . The term  $(\lambda x : \mathbb{A}.v)$  is the lambda abstraction term, which represents a function that takes an argument of type  $\mathbb{A}$  and returns the value of  $v$ . The term  $(vw)$  is the application term, which applies the function  $v$  to the argument  $w$ . The term  $v \otimes w$  is the tensor product of  $v$  and  $w$ . The term  $\text{pm } v \text{ to } x \otimes y.w$  is the pattern-matching construct, which is used to deconstruct a tensor product into components  $x$  and  $y$ . The term  $v \text{ to } *.w$  is used to discard a variable  $v$  of the unit type. The term  $\text{dis}(v)$  is the discard term, which is used to discard a term  $v$ . The terms  $\text{inl}_{\mathbb{B}}(v)$  and  $\text{inr}_{\mathbb{A}}(v)$  represent the left and right injections of  $v$ , respectively. Intuitively, the case statement executes  $w$  when  $v$  is a left injection, and  $u$  when  $v$  is a right injection, and a “mixture” of both otherwise.

Ver o que por antes do ::= porque v1,..., vn tb são termos

### 2.2.3 Free and Bound Variables

An occurrence of a variable  $x$  within a term of the form  $\lambda x.v$  is referred to as *bound*. Similarly, the variables  $x$  and  $y$  in the term  $\text{pm } v \text{ to } x \otimes y.w$  are also bound. A variable occurrence that is not bound is said to be *free*. For example, in the term  $\lambda x.xy$ , the variable  $y$  is free, whereas the variable  $x$  is bound.

The set of free variables of a term  $v$  is denoted by  $FV(v)$ , and is defined inductively as follows:

$$\begin{aligned}
FV(x) &= \{x\}, & FV(*) &= \emptyset, \\
FV(f(v_1, \dots, v_n)) &= FV(v_1) \cup \dots \cup FV(v_n) & FV(\lambda x : \mathbb{A}.v) &= FV(v) \setminus \{x\}, \\
FV(vw) &= FV(v) \cup FV(w), & FV(v \otimes w) &= FV(v) \cup FV(w), \\
FV(\text{pm } v \text{ to } x \otimes y.w) &= FV(v) \cup (FV(w) \setminus \{x, y\}) & FV(\text{dis}(v)) &= FV(v), \\
FV(v \text{ to } * .w) &= FV(v) \cup FV(w) & FV(\text{inl}_{\mathbb{B}}(v)) &= FV(\text{inr}_{\mathbb{A}}(v)) = FV(v). \\
FV(\text{case } v \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow w; \text{inr}_{\mathbb{A}}(y) \Rightarrow u \}) &= FV(v) \cup (FV(w) \setminus \{x\}) \cup (FV(u) \setminus \{y\})
\end{aligned}$$

## 2.2.4 Term formation rules

To prevent the formation of nonsensical terms within the context of lambda calculus, such as  $(v \otimes w)(u)$ , the *typing rules* are imposed.

A *typed term* is a pair consisting of a term and its corresponding type. The notation  $v : \mathbb{A}$  denotes that the term  $v$  has type  $\mathbb{A}$ . Typing rules are formulated using *typing judgments*. A typing judgment is an expression of the form  $x_1 : \mathbb{A}_1, \dots, x_n : \mathbb{A}_n \triangleright v : \mathbb{A}$  (where  $n \geq 1$ ), which asserts that the term  $v$  is a well-typed term of type  $\mathbb{A}$  under the assumption that each variable  $x_i$  has type  $\mathbb{A}_i$ , for  $1 \leq i \leq n$ . The list  $x_1 : \mathbb{A}_1, \dots, x_n : \mathbb{A}_n$  of typed variables is called the *typing context* of the judgment, and it might be empty. Each variable  $x_i$  (where  $1 \leq i \leq n$ ) must occur at most once in  $x_1, \dots, x_n$ . The typing contexts are denoted by Greek letters  $\Gamma, \Delta, E$ , and from now on, when referring to an abstract judgment, the notation  $\Gamma \triangleright v : \mathbb{A}$  will be employed. The empty context is denoted by  $-$ . Note that in the affine lambda calculus, different contexts do not share variables. For example, if  $\Gamma = x : \mathbb{A}, y : \mathbb{B}$  none of these variables can appear in any other context.

The concept of *shuffling* is employed to construct a linear typing system that ensures the admissibility of the exchange rule and enables unambiguous reference to judgment's denotation  $\llbracket \Gamma \triangleright v : \mathbb{A} \rrbracket$ . An admissible rule is not explicitly included in the formal definition of type theory, but its validity can be proven by demonstrating that whenever the premises can be derived, it is possible to construct a derivation of its conclusion. Shuffling is defined as a permutation of typed variables in a sequence of contexts,  $\Gamma_1, \dots, \Gamma_n$ , preserving the relative order of variables within each  $\Gamma_i$  [36]. For instance, if  $\Gamma_1 = x : \mathbb{A}, y : \mathbb{B}$  and  $\Gamma_2 = z : \mathbb{D}$ , then  $z : \mathbb{D}, x : \mathbb{A}, y : \mathbb{B}$  is a valid shuffle of  $\Gamma_1, \Gamma_2$ . On the other hand,  $y : \mathbb{B}, x : \mathbb{A}, z : \mathbb{D}$  is not a shuffle because it alters the occurrence order of  $x$  and  $y$  in  $\Gamma_1$ . The set of shuffles

in  $\Gamma_1, \dots, \Gamma_n$  is denoted as  $\text{Sf}(\Gamma_1, \dots, \Gamma_n)$ . A valid typing derivation is constructed using the inductive rules shown in Figure 1.

$$\begin{array}{c}
\frac{\Gamma_i \triangleright v_i : \mathbb{A}_i \quad f : \mathbb{A}_1, \dots, \mathbb{A}_n \rightarrow \mathbb{A} \in \Sigma \quad E \in \text{Sf}(\Gamma_1; \dots; \Gamma_n)}{E \triangleright f(v_1, \dots, v_n) : \mathbb{A}} (\text{ax}) \quad \frac{}{x : \mathbb{A} \triangleright x : \mathbb{A}} (\text{hyp}) \\
\\
\frac{}{- \triangleright * : \mathbb{I}} (\mathbb{I}_i) \quad \frac{\Gamma \triangleright v : \mathbb{A} \otimes \mathbb{B} \quad \Delta, x : \mathbb{A}, y : \mathbb{B} \triangleright w : \mathbb{D} \quad E \in \text{Sf}(\Gamma; \Delta)}{E \triangleright \text{pm } v \text{ to } x \otimes y.w : \mathbb{D}} (\otimes_e) \\
\\
\frac{\Gamma \triangleright v : \mathbb{A} \quad \Delta \triangleright w : \mathbb{B} \quad E \in \text{Sf}(\Gamma; \Delta)}{E \triangleright v \otimes w : \mathbb{A} \otimes \mathbb{B}} (\otimes_i) \quad \frac{\Gamma \triangleright v : \mathbb{I} \quad \Delta \triangleright w : \mathbb{A} \quad E \in \text{Sf}(\Gamma; \Delta)}{E \triangleright v \text{ to } *.w : \mathbb{A}} (\mathbb{I}_e) \\
\\
\frac{\Gamma, x : \mathbb{A} \triangleright v : \mathbb{B}}{\Gamma \triangleright \lambda x : \mathbb{A}.v : \mathbb{A} \multimap \mathbb{B}} (-\circ_i) \quad \frac{\Gamma \triangleright v : \mathbb{A} \multimap \mathbb{B} \quad \Delta \triangleright w : \mathbb{A} \quad E \in \text{Sf}(\Gamma; \Delta)}{E \triangleright vw : \mathbb{B}} (-\circ_e) \\
\\
\frac{\Gamma \triangleright v : \mathbb{A}}{\Gamma \triangleright \text{dis}(v) : \mathbb{I}} (\text{dis}) \quad \frac{\Gamma \triangleright v : \mathbb{A}}{\Gamma \triangleright \text{inl}_{\mathbb{B}}(v) : \mathbb{A} \oplus \mathbb{B}} (\text{inl}) \quad \frac{\Gamma \triangleright v : \mathbb{B}}{\Gamma \triangleright \text{inr}_{\mathbb{A}}(v) : \mathbb{A} \oplus \mathbb{B}} (\text{inr}) \\
\\
\frac{\Gamma \triangleright v : \mathbb{A} \oplus \mathbb{B} \quad \Delta, x : \mathbb{A} \triangleright w : \mathbb{D} \quad \Delta, y : \mathbb{B} \triangleright u : \mathbb{D} \quad E \in \text{Sf}(\Gamma; \Delta)}{E \triangleright \text{case } v \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow w; \text{inr}_{\mathbb{A}}(y) \Rightarrow u \} : \mathbb{D}} (\text{case})
\end{array}$$

Figure 1: Term formation rules of affine lambda calculus.

The rule (ax) states that if there is a function  $f \in \Sigma$  that has type  $\mathbb{A}_1, \dots, \mathbb{A}_n \rightarrow \mathbb{A}$  and a set of variables  $v_1, \dots, v_n$  whose types match the type of the arguments of  $f$ , then if that function is applied to  $v_1, \dots, v_n$  the respective result is of type  $\mathbb{A}$ . The rule (hyp) is a tautology: under the assumption that  $x$  has type  $\mathbb{A}$ ,  $x$  has type  $\mathbb{A}$ . The rule ( $\mathbb{I}_i$ ) asserts that the unit element  $*$  always has type  $\mathbb{I}$ . The rule ( $-\circ_i$ ) expresses that if  $v$  is a term of type  $\mathbb{B}$  with a variable  $x$  of type  $\mathbb{A}$ , then  $\lambda x : \mathbb{A}.v$  is a function of type  $\mathbb{A} \multimap \mathbb{B}$ . The rule ( $-\circ_e$ ) states that a function of type  $\mathbb{A} \multimap \mathbb{B}$  can be applied to an argument of type  $\mathbb{A}$  to produce a result of type  $\mathbb{B}$ . The rule ( $\otimes_i$ ) asserts that if there is a term  $v$  of type  $\mathbb{A}$  and a term  $w$  of type  $\mathbb{B}$ , then the tensor of these terms is of type  $\mathbb{A} \otimes \mathbb{B}$ . The rule ( $\otimes_e$ ) expresses if there is a term  $w$  of type  $\mathbb{D}$  with variables  $x$  and  $y$  of types  $\mathbb{A}$  and  $\mathbb{B}$ , respectively, and a term  $v$  of type  $\mathbb{A} \otimes \mathbb{B}$ , then  $v$  can be deconstructed into  $x \otimes y$ . The rule ( $\mathbb{I}_e$ ) states that if there is a term  $w$  of type  $\mathbb{A}$  and a term  $v$  of type  $\mathbb{I}$ , then  $v$  can be discarded, and only the term  $w$  remains. The rule (dis) asserts that a term  $v$  of type  $\mathbb{A}$  can be discarded, resulting in a term of type  $\mathbb{I}$ . The rules ( $\text{inl}_{\mathbb{B}}$ ) (resp. ( $\text{inr}_{\mathbb{A}}$ )) states that if we inject a term  $v$  of type  $\mathbb{A}$  (resp.  $\mathbb{B}$ ) into  $\mathbb{A} \oplus \mathbb{B}$  we obtain a term of type  $\mathbb{A} \oplus \mathbb{B}$ . Finally, the rule

(case) states that if there are two programs of type  $\mathbb{D}$  to be executed depending on the value of a term  $v$  of type  $\mathbb{A} \oplus \mathbb{B}$  (whose right and left components are assigned to variables  $x$  and  $y$ , respectively), then the resulting program also has type  $\mathbb{D}$ .

For a better understanding of the rules, a few straightforward programming examples are provided.

**Example 2.2.1.** For instance, the program that swaps the elements of a tensor product can be written as follows:

$$\mathbf{SwapTensor} \triangleq x : \mathbb{A} \otimes \mathbb{B} \triangleright \mathbf{pm} \ x \ \mathbf{to} \ a \otimes b. b \otimes a : \mathbb{B} \otimes \mathbb{A}$$

Now, to prove that this program is well-typed one can write the following typing derivation:

$$\begin{array}{ll} 1 \quad x : \mathbb{A} \otimes \mathbb{B} \triangleright y : \mathbb{A} \otimes \mathbb{B} & (\text{hyp}) \\ 2 \quad b : \mathbb{B} \triangleright b : \mathbb{B} & (\text{hyp}) \\ 3 \quad a : \mathbb{A} \triangleright a : \mathbb{A} & (\text{hyp}) \\ 4 \quad b : \mathbb{B}, a : \mathbb{A} \triangleright b \otimes a : \mathbb{B} \otimes \mathbb{A} & (2, 3, \otimes_i) \\ 5 \quad x : \mathbb{A}, y : \mathbb{B} \triangleright \mathbf{pm} \ x \otimes y \ \mathbf{to} \ a \otimes b. b \otimes a : \mathbb{B} \otimes \mathbb{A} & (1, 4, \otimes_e) \end{array}$$

Observe that in the notation of the third column, the numbers correspond to the premises utilized in the application of the rule.

**Example 2.2.2.** Another example is the function that receives a tensor product and returns first element, discarding the second:

$$\mathbf{Dis2nd} \triangleq - \triangleright \lambda x : \mathbb{A} \otimes \mathbb{A}. \mathbf{pm} \ x \ \mathbf{to} \ a \otimes b. \mathbf{dis}(b) \ \mathbf{to} \ *.a : \mathbb{A}$$

To prove that this program is well-typed one can write the following typing derivation:

$$\begin{array}{ll} 1 \quad b : \mathbb{A} \triangleright b : \mathbb{A} & (\text{hyp}) \\ 2 \quad b : \mathbb{A} \triangleright \mathbf{dis}(b) : \mathbb{I} & (1, \mathbf{dis}) \\ 3 \quad a : \mathbb{A} \triangleright a : \mathbb{A} & (\text{hyp}) \\ 4 \quad a : \mathbb{A}, b : \mathbb{A} \triangleright \mathbf{dis}(b) \ \mathbf{to} \ *.a & (2, 3, \mathbb{I}_e) \\ 5 \quad x : \mathbb{A} \otimes \mathbb{A} \triangleright x : \mathbb{A} \otimes \mathbb{A} & (\text{hyp}) \\ 6 \quad x : \mathbb{A} \otimes \mathbb{A} \triangleright \mathbf{pm} \ x \ \mathbf{to} \ a \otimes b. \mathbf{dis}(b) \ \mathbf{to} \ *.a : \mathbb{A} & (4, 5, \otimes_e) \\ 7 \quad - \triangleright \lambda x : \mathbb{A} \otimes \mathbb{A}. \mathbf{pm} \ x \ \mathbf{to} \ a \otimes b. \mathbf{dis}(b) \ \mathbf{to} \ *.a : \mathbb{A} & (6, -\circ_i) \end{array}$$

**Example 2.2.3.** Consider an analogous program to the previous example that receives a tensor product and returns second element, discarding the first:

$$\mathbf{Dis1st} \triangleq - \triangleright \lambda x : \mathbb{A} \otimes \mathbb{A}. \text{pm } x \text{ to } a \otimes b. \text{dis}(a) \text{ to } *. b : \mathbb{A} \otimes \mathbb{A} \multimap \mathbb{A}$$

Following a similar line of reasoning as in the previous example, it is straightforward to verify that this  $\lambda$ -term is also well-typed. Now, let us examine a program that receives both a co-product/sum and a tensor product, and applies either **Dis1st** or **Dis2nd** to the tensor product (or a mixture of both), depending on the value of the coproduct:

$$\mathbf{Dis1stOR2nd} \triangleq - \triangleright \lambda x : \mathbb{B} \oplus \mathbb{B}. \lambda y : \mathbb{A} \otimes \mathbb{A}. \text{case } x \left\{ \begin{array}{l} \text{inl}_{\mathbb{B}}(w) \Rightarrow \text{dis}(w) \text{ to } *. \mathbf{Dis1st } y; \\ \text{inr}_{\mathbb{B}}(z) \Rightarrow \text{dis}(z) \text{ to } *. \mathbf{Dis2nd } y \end{array} \right\}$$

To prove that this program is well-typed one can write the following typing derivation (the derivations regarding **Dis1st** and **Dis2nd** will be omitted):

$$\begin{array}{ll}
1 & y : \mathbb{A} \otimes \mathbb{A} \triangleright y : \mathbb{A} \otimes \mathbb{A} \quad (\text{hyp}) \\
2 & y : \mathbb{A} \otimes \mathbb{A} \triangleright \mathbf{Dis1st } y : \mathbb{A} \quad (-\circ_e) \\
3 & y : \mathbb{A} \otimes \mathbb{A} \triangleright \mathbf{Dis2nd } y : \mathbb{A} \quad (-\circ_e) \\
4 & w : \mathbb{B} \triangleright w : \mathbb{B} \quad (\text{hyp}) \\
5 & w : \mathbb{B} \triangleright \text{dis}(w) : \mathbb{I} \quad (4, \text{dis}) \\
6 & z : \mathbb{B} \triangleright z : \mathbb{B} \quad (\text{hyp}) \\
7 & z : \mathbb{B} \triangleright \text{dis}(z) : \mathbb{I} \quad (6, \text{dis}) \\
8 & y : \mathbb{A} \otimes \mathbb{A}, w : \mathbb{B} \triangleright \text{dis}(w) \text{ to } *. \mathbf{Dis1st } y : \mathbb{A} \quad (2, 5, \mathbb{I}_e) \\
9 & y : \mathbb{A} \otimes \mathbb{A}, z : \mathbb{B} \triangleright \text{dis}(z) \text{ to } *. \mathbf{Dis2nd } y : \mathbb{A} \quad (3, 7, \mathbb{I}_e) \\
10 & x : \mathbb{B} \oplus \mathbb{B} \triangleright x : \mathbb{B} \oplus \mathbb{B} \quad (\text{hyp}) \\
11 & x : \mathbb{B} \oplus \mathbb{B}, y : \mathbb{A} \otimes \mathbb{A} \triangleright \text{case } x \left\{ \begin{array}{l} \text{inl}_{\mathbb{B}}(w) \Rightarrow \text{dis}(w) \text{ to } *. \mathbf{Dis1st } y; \\ \text{inr}_{\mathbb{B}}(z) \Rightarrow \text{dis}(z) \text{ to } *. \mathbf{Dis2nd } y \end{array} \right\} \quad (8, 9, 10, \text{case}) \\
12 & x : \mathbb{B} \oplus \mathbb{B} \triangleright \lambda y : \mathbb{A} \otimes \mathbb{A}. \text{case } x \{ \dots \} \quad (11, -\circ_i) \\
13 & - \triangleright \lambda x : \mathbb{B} \oplus \mathbb{B}. \lambda y : \mathbb{A} \otimes \mathbb{A}. \text{case } x \{ \dots \} \quad (12, -\circ_i)
\end{array}$$

## 2.2.5 $\alpha$ -equivalence

A natural notion of equivalence definition stems from the fact that terms that differ only in the names of their bound variables represent the same program. For instance, the functions



$\lambda x : \mathbb{A}.x$  and  $\lambda y : \mathbb{A}.y$  have the same input-output behavior, despite being represented by different lambda terms. This equivalence is called  $\alpha$ -equivalence.

**Definition 2.2.4.** The  $\alpha$ -equivalence is an equivalence relation on lambda terms that is used to rename bound variables. To rename a variable  $x$  as  $y$  in a term  $v$ , denoted by  $v\{y/x\}$ , is to replace all occurrences of  $x$  in  $v$  by  $y$ . Two terms  $v$  and  $w$  are  $\alpha$ -equivalent, written  $=_\alpha$ , if one can be derived from the other by a series of changes of bound variables

**Convention 2.2.5.** Terms are considered up to  $\alpha$ -equivalence from now on.

## 2.2.6 Substitution

The substitution of a variable  $x$  for a term  $w$  in a term  $v$  is denoted by  $v[w/x]$ . It is only permitted to replace free variables. For instance,  $\lambda x.x[v/x]$  is  $\lambda x.x$  and not  $\lambda x.v$ . Moreover, it is necessary to avoid the unintended binding of free variables. For example,

$$(\text{pm } x \otimes y \text{ to } a \otimes b.b \otimes a \otimes z) [z/\text{pm } c \otimes d \text{ to } e \otimes f.f \otimes e \otimes a]$$

is not the same as

$$\text{pm } x \otimes y \text{ to } a \otimes b.b \otimes a \otimes (\text{pm } c \otimes d \text{ to } e \otimes f.f \otimes e \otimes a).$$

Instead, the bounded variable  $a$  must be renamed before the substitution, and in this case, the proper substitution is

$$(\text{pm } x \otimes y \text{ to } t \otimes b.b \otimes t \otimes z) [z/\text{pm } c \otimes d \text{ to } e \otimes f.f \otimes e \otimes a]$$

which is equal to

$$\text{pm } x \otimes y \text{ to } t \otimes b.b \otimes t \otimes (\text{pm } c \otimes d \text{ to } e \otimes f.f \otimes e \otimes a).$$

Note that a simple way of ensuring these restrictions are satisfied is not allowing the variable  $x$  to occur in the context of  $w$  in  $v[w/x]$ . Since  $x$  is in the context of  $v$ , this is always the case in the affine lambda calculus.

**Definition 2.2.6.** Given the typings judgments  $\Gamma, x : \mathbb{A} \triangleright v : \mathbb{B}$  and  $\Delta \triangleright w : \mathbb{A}$ , the substitution  $\Gamma, \Delta \triangleright v[w/x] : \mathbb{B}$  is defined below. The types of judgments are omitted as no ambiguity

arises.

$$\Gamma, \Delta \triangleright y[w/x] = \Gamma, \Delta \triangleright y,$$

$$\Delta \triangleright *[w/x] = \Delta \triangleright *,$$

$$\Gamma, \Delta \triangleright (\lambda y : \mathbb{B}.v)[w/x] = \Gamma, \Delta \triangleright \lambda y : \mathbb{B}.v[w/x],$$

$$(\mathbf{dis}(v))[w/x] = \mathbf{dis}(v[w/x]),$$

$$(\mathbf{inl}(v))[w/x] = \mathbf{inl}(v[w/x]),$$

$$(\mathbf{inr}(v))[w/x] = \mathbf{inr}(v[w/x]),$$

In the next three cases,  $\Gamma, x : \mathbb{A} \in \mathbf{Sf}(\Gamma_1, \dots, \Gamma_i, \dots, \Gamma_n)$  and  $\Gamma_i \triangleright v_i$

$$\Gamma, \Delta \triangleright (f(v_1, \dots, v_n))[w/x] = \Gamma, \Delta \triangleright f(v_1[w/x], \dots, v_n), \quad (\text{if } x : \mathbb{A} \in \Gamma_1)$$

$$\Gamma, \Delta \triangleright (f(v_1, \dots, v_i, \dots, v_n))[w/x] = \Gamma, \Delta \triangleright f(v_1, \dots, v_i[w/x], \dots, v_n), \quad (\text{if } x : \mathbb{A} \in T_i)$$

$$\Gamma, \Delta \triangleright (f(v_1, \dots, v_n))[w/x] = \Gamma, \Delta \triangleright f(v_1, \dots, v_n[w/x]), \quad (\text{if } x : \mathbb{A} \in \Gamma_n)$$

In the next two cases,  $\Gamma, x : \mathbb{A} \in \mathbf{Sf}(\Gamma_1, \Gamma_2), \Gamma_1 \triangleright v$ , and  $\Gamma_2 \triangleright u$

$$\Gamma, \Delta \triangleright (vu)[w/x] = \Gamma, \Delta \triangleright (v[w/x]u), \quad (\text{if } x : \mathbb{A} \in \Gamma_1)$$

$$\Gamma, \Delta \triangleright (vu)[w/x] = \Gamma, \Delta \triangleright (vu[w/x]), \quad (\text{if } x : \mathbb{A} \in \Gamma_2)$$

In the next two cases,  $\Gamma, x : \mathbb{A} \in \mathbf{Sf}(\Gamma_1, \Gamma_2), \Gamma_1 \triangleright v$ , and  $\Gamma_2 \triangleright u$

$$\Gamma, \Delta \triangleright (v \otimes u)[w/x] = \Gamma, \Delta \triangleright v[w/x] \otimes u, \quad (\text{if } x : \mathbb{A} \in \Gamma_1)$$

$$\Gamma, \Delta \triangleright (v \otimes u)[w/x] = \Gamma, \Delta \triangleright v \otimes u[w/x], \quad (\text{if } x : \mathbb{A} \in \Gamma_2)$$

In the next two cases,  $\Gamma, x : \mathbb{A} \in \mathbf{Sf}(\Gamma_1, \Gamma_2), \Gamma_1 \triangleright v$ , and  $\Gamma_2, y : \mathbb{D}, z : \mathbb{E} \triangleright u$

$$\Gamma, \Delta \triangleright (\mathbf{pm} \ v \ \mathbf{to} \ y \otimes z.u)[w/x] = \Gamma, \Delta \triangleright \mathbf{pm} \ v[w/x] \ \mathbf{to} \ y \otimes z.u, \quad (\text{if } x : \mathbb{A} \in \Gamma_1)$$

$$\Gamma, \Delta \triangleright (\mathbf{pm} \ v \ \mathbf{to} \ y \otimes z.u)[w/x] = \Gamma, \Delta \triangleright \mathbf{pm} \ v \ \mathbf{to} \ y \otimes z.u[w/x], \quad (\text{if } x : \mathbb{A} \in \Gamma_2)$$

In the next two cases,  $\Gamma, x : \mathbb{A} \in \mathbf{Sf}(\Gamma_1, \Gamma_2), \Gamma_1 \triangleright v$ , and  $\Gamma_2 \triangleright u$

$$\Gamma, \Delta \triangleright (v \ \mathbf{to} \ * .u)[w/x] = \Gamma, \Delta \triangleright v[w/x] \ \mathbf{to} \ * .u \quad (\text{if } x : \mathbb{A} \in \Gamma_1),$$

$$\Gamma, \Delta \triangleright (v \ \mathbf{to} \ * .u)[w/x] = \Gamma, \Delta \triangleright v \ \mathbf{to} \ * .u[w/x] \quad (\text{if } x : \mathbb{A} \in \Gamma_2).$$

In the next two cases,  $\Gamma, x : \mathbb{A} \in \mathbf{Sf}(\Gamma_1, \Gamma_2), \Gamma_1 \triangleright v, \Gamma_2, y : \mathbb{D} \triangleright a$ , and

$$\Gamma_2, z : \mathbb{E} \triangleright b$$

$$\mathbf{case} \ v \ \left\{ \begin{array}{l} \mathbf{inl}_{\mathbb{I}}(y) \Rightarrow a; \\ \mathbf{inr}_{\mathbb{I}}(z) \Rightarrow b \end{array} \right\} [w/x] = \mathbf{case} \ v[w/x] \ \left\{ \begin{array}{l} \mathbf{inl}_{\mathbb{I}}(y) \Rightarrow a; \\ \mathbf{inr}_{\mathbb{I}}(z) \Rightarrow b \end{array} \right\} \quad (\text{if } x : \mathbb{A} \in \Gamma_1)$$

$$\text{case } v \left\{ \begin{array}{l} \text{inl}_{\mathbb{E}}(y) \Rightarrow a; \\ \text{inr}_{\mathbb{D}}(z) \Rightarrow b \end{array} \right\} [w/x] = \text{case } v \left\{ \begin{array}{l} \text{inl}_{\mathbb{E}}(y) \Rightarrow a[w/x]; \\ \text{inr}_{\mathbb{D}}(z) \Rightarrow b[w/x] \end{array} \right\} \quad (\text{if } x : \mathbb{A} \in \Gamma_2)$$

The sequential substitutions  $M[M_i/x_i] \dots [M_n/x_n]$  are written as  $M[M_i/x_i, \dots, M_n/x_n]$ .

### 2.2.7 Properties

The calculus defined in Figure 1 possesses several desirable properties, which are listed below. Before proceeding, it is necessary to introduce some auxiliary notation. Given a context  $\Gamma$ ,  $te(\Gamma)$  denotes context  $\Gamma$  with all types erased. The expression  $\Gamma \simeq_{\pi} \Gamma'$  denotes that the contexts  $\Gamma$  and  $\Gamma'$  are a permutation of each other. This notation also applies to non-repetitive lists of untyped variables  $te(\Gamma)$ . Additionally, a judgment  $\Gamma \triangleright v : \mathbb{A}$  will often be abbreviated into  $\Gamma \triangleright v$  or even just  $v$  when no ambiguities arise.

**Theorem 2.2.7.** *The lambda calculus defined by the rules of Figure 1 has the following properties:*

1. *for all judgements  $\Gamma \triangleright v$  and  $\Gamma' \triangleright v$ ,  $te(\Gamma) \simeq_{\pi} te(\Gamma')$ ;*
2. *additionally if  $\Gamma \triangleright v : \mathbb{A}$ ,  $\Gamma' \triangleright v : \mathbb{A}'$ , and  $\Gamma \simeq_{\pi} \Gamma'$ , then  $\mathbb{A}$  must be equal to  $\mathbb{A}'$ ;*
3. *all judgements  $\Gamma \triangleright v : \mathbb{A}$  have a unique derivation.*

*Proof.* Since these properties are established in [37, Theorem 2.3] for the lambda calculus without conditionals, it suffices to consider the cases involving conditionals.

It follows in all three cases from induction over the length of derivation trees.

Let us focus first on Property (1). The case of the rules concerning injections is direct. As for rule (case) take two contexts  $E$  and  $E'$  for the same conditional. According to this rule we obtain contexts  $\Gamma, \Gamma', \Delta, \Delta'$  such that  $E \in \text{Sf}(\Gamma; \Delta)$  and  $E' \in \text{Sf}(\Gamma'; \Delta')$ . It follows from induction that  $\Gamma \simeq_{\pi} \Gamma'$  and  $\Delta \simeq_{\pi} \Delta'$ , and the proof is then obtained from the sequence of equivalences,

$$\begin{aligned} te(E) &\simeq_{\pi} te(\Gamma, \Delta) \\ &\simeq_{\pi} te(\Gamma', \Delta') \\ &\simeq_{\pi} te(E') \end{aligned}$$

Concerning Property (2), the case of the rules concerning injections is direct and the case of rule (case) is a corollary of Property (1). Finally let us consider Property (3). Again the case concerning injections is direct and we thus focus only on rule (case). According to this rule we obtain contexts  $\Gamma, \Gamma', \Delta, \Delta'$  such that  $E \in \text{Sf}(\Gamma; \Delta)$  and  $E \in \text{Sf}(\Gamma'; \Delta')$ . By an appeal to Property (1) we also obtain  $\Gamma \simeq_\pi \Gamma'$  and  $\Delta \simeq_\pi \Delta'$ , and thus since shuffling preserves relative orders we obtain  $\Gamma = \Gamma'$  and  $\Delta = \Delta'$ . The proof then follows by induction.  $\square$

**Lemma 2.2.8** (Exchange and Substitution). *For every judgement  $\Gamma, x : \mathbb{A}, y : \mathbb{B}, \Delta \triangleright v : \mathbb{D}$  the judgement  $\Gamma, y : \mathbb{B}, x : \mathbb{A}, \Delta \triangleright v : \mathbb{D}$  is derivable. Not only this, given judgements  $\Gamma, x : \mathbb{A} \triangleright v : \mathbb{B}$  and  $\Delta \triangleright w : \mathbb{A}$  the judgement  $\Gamma, \Delta \triangleright v[w/x] : \mathbb{B}$  is also derivable.*

*Proof.* Once again, these properties are established in [24, Theorem 2.1] for the lambda calculus without conditionals, so it suffices to consider the cases involving conditionals.

We start with the exchange property which follows by induction over the length of derivation trees. The rules that involve injections are direct. The rule (case) calls for case distinction, more specifically we distinguish between the cases in which both variables ( $x : \mathbb{A}$  and  $y : \mathbb{B}$ ) are in  $\Gamma$ , both are in  $\Delta$ , and otherwise. The first two cases follow straightforwardly by induction and the definition of a shuffle. For the third case consider a judgement  $E_1, x : \mathbb{A}, y : \mathbb{B}, E_2 \triangleright \text{case } v \{ \text{inl}_{\mathbb{F}}(a) \Rightarrow w; \text{inr}_{\mathbb{E}}(b) \Rightarrow u \} : \mathbb{D}$ , and assume with no loss of generality that  $\Gamma$  is of the form  $\Gamma_1, x : \mathbb{A}, \Gamma_2$  and  $\Delta$  of the form  $\Delta_1, y : \mathbb{B}, \Delta_2$ . The proof now follows directly from the implication,

$$\begin{aligned} & E_1, x : \mathbb{A}, y : \mathbb{B}, E_2 \in \text{Sf}(\Gamma_1, x : \mathbb{A}, \Gamma_2; \Delta_1, y : \mathbb{B}, \Delta_2) \\ \implies & E_1, y : \mathbb{B}, x : \mathbb{A}, E_2 \in \text{Sf}(\Gamma_1, x : \mathbb{A}, \Gamma_2; \Delta_1, y : \mathbb{B}, \Delta_2) \end{aligned}$$

(which holds by the definition of a shuffle).

Finally we now focus on the substitution rule which also follows by induction over the length of derivation trees. Again the cases involving the injections are direct, and we thus only detail the proof of rule (case). Consider then judgements  $E, x : \mathbb{A} \triangleright \text{case } v \{ \text{inl}_{\mathbb{D}}(a) \Rightarrow w; \text{inr}_{\mathbb{E}}(b) \Rightarrow u \} : \mathbb{B}$  and  $Z \triangleright t : \mathbb{A}$  with  $E \in \text{Sf}(\Gamma; \Delta)$ . According to the definition of a shuffle either  $\Gamma$  is of the form  $\Gamma_1, x : \mathbb{A}$  or  $\Delta$  is of the form  $\Delta_1, x : \mathbb{A}$ . The first case follows directly and the second case is a corollary of the exchange rule.  $\square$

## 2.2.8 Equations-in-context

The simply typed lambda calculus is a formal language that captures operations like the application of a function to an argument and the elimination of variables. To express these operations there is a set of rules known as reduction rules. These rules fall into two primary categories: the  $\beta$ -reductions, which perform operations and enforce the implicit meaning of the term, and  $\eta$ -reductions, which simplify terms by exploiting the extensionality of functions. There is also a secondary class of reductions known as *commuting conversions*, which serve to disambiguate terms that, while equivalent, have different representations. As a result, affine  $\lambda$ -calculus comes equipped with the so-called equations-in-context  $\Gamma \triangleright v = w : \mathbb{A}$ , depicted in Figure 2.

$(\beta)$	$\Gamma, \Delta \triangleright (\lambda x : \mathbb{A}. v) w = v[w/x] : \mathbb{B}$	$(\eta)$	$\Gamma \triangleright \lambda x : \mathbb{A}. (v x) = v : \mathbb{A} \multimap \mathbb{B}$
$(\beta_{\mathbb{I}_e})$	$\Gamma \triangleright * \text{ to } * . v = v : \mathbb{A}$	$(\eta_{\mathbb{I}_e})$	$\Delta, \Gamma \triangleright v \text{ to } * . w[* / z] = w[v / z] : \mathbb{A}$
$(\beta_{\otimes_e})$	$E, \Gamma, \Delta \triangleright \text{pm } v \otimes w \text{ to } x \otimes y. u = u[v / x, w / y] : \mathbb{A}$		
$(\eta_{\otimes_e})$	$\Delta, \Gamma \triangleright \text{pm } v \text{ to } x \otimes y. u[x \otimes y / z] = u[v / z] : \mathbb{A}$		
$(c_{\mathbb{I}_e})$	$\Delta, \Gamma, E \triangleright u[v \text{ to } * . w / z] = v \text{ to } * . u[w / z] : \mathbb{A}$		
$(c_{\otimes_e})$	$\Delta, \Gamma, E \triangleright u[\text{pm } v \text{ to } x \otimes y. w / z] = \text{pm } v \text{ to } x \otimes y. u[w / z] : \mathbb{A}$		
$(\eta_{\text{dis}})$	$x_1 : \mathbb{A}_1, \dots, x_n : \mathbb{A}_n \triangleright v = \text{dis}(x_1) \text{ to } * \dots \text{dis}(x_{n-1}) \text{ to } * \text{dis}(x_n) : \mathbb{I}$		
$(\beta_{\text{case}}^{\text{inl}})$	$\text{case inl}_{\mathbb{B}}(v) \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow w; \text{inr}_{\mathbb{A}}(y) \Rightarrow u \} = w[v / x]$		
$(\beta_{\text{case}}^{\text{inr}})$	$\text{case inl}_{\mathbb{B}}(v) \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow w; \text{inr}_{\mathbb{A}}(y) \Rightarrow u \} = u[v / y]$		
$(\eta_{\text{case}})$	$\text{case } v \{ \text{inl}_{\mathbb{B}}(y) \Rightarrow w[\text{inl}_{\mathbb{B}}(y) / x]; \text{inr}_{\mathbb{A}}(z) \Rightarrow w[\text{inr}_{\mathbb{A}}(z) / x] \} = w[v / x]$		

Figure 2: Equations-in-context for affine lambda calculus

It is evident that, for example, equation  $(\beta)$  enforces the meaning of  $(\lambda x : \mathbb{A}. v) w$ , which is interpreted as “ $v$  with  $w$  in place of  $x$ ”. The equation  $(\eta)$ , on the other hand, is a simplification rule that states that a function that applies another function  $v$  to an argument  $x$  can be simplified to the function  $v$  itself. The remaining  $\beta$  e  $\eta$  equations follow similar reasoning. The commuting conversion  $(c_{\mathbb{I}_e})$  expresses that substituting a variable  $z$  by a term that maps a term  $v$  to the unit element  $*$  in a term  $w$  is equivalent to mapping a term  $v$  to the unit element  $*$  and then replacing  $z$  by  $w$ . The other commuting conversion has a similar interpretation.

**Example 2.2.9.** For instance, consider the  $\lambda$ -term

$$\multimap \triangleright (\lambda z : \mathbb{I} \otimes \mathbb{A}. \text{pm } z \text{ to } * \otimes y. y) (v \otimes w) : \mathbb{A}$$

Applying the  $\beta$  reduction, we have:

$$-\triangleright (\lambda z : \mathbb{I} \otimes \mathbb{A}. \text{pm } z \text{ to } * \otimes y. y) (v \otimes w) = \text{pm } v \otimes w \text{ to } * \otimes y. y : \mathbb{A}.$$

Next, applying the  $\beta_{\otimes_e}$  reduction, it follows:

$$\text{pm } v \otimes w \text{ to } * \otimes y. y : \mathbb{A} = w : \mathbb{A}.$$

Não sei se este é o lugar mais indicado para estas definições, mas tb não estou a ver outro dado que as categorias só vêm a seguir. Ou então passar isto para a interpretação.

**Definition 2.2.10.** Let  $S$  be a set. A *relation* on  $S$  is a subset  $R \subseteq S \times S$ . An ordered pair  $(s_1, s_2) \in R$  means that  $s_1$  is related to  $s_2$ .

**Definition 2.2.11.** A relation on a set  $S$  is an *equivalence relation* if it is

- reflexive: for all  $x \in S$ ,  $x \sim x$ ,
- symmetric: for all  $x, y \in S$ , if  $x \sim y$  then  $y \sim x$ , and
- transitive: for all  $x, y, z \in S$ , if  $x \sim y$  and  $y \sim z$ , then  $x \sim z$ .

We denote such a relation by  $\sim \subseteq S \times S$ , and write  $x \sim y$  to mean that  $(x, y) \in \sim$ .

**Definition 2.2.12.** Given an equivalence relation on a set  $S$ , we can describe disjoint subsets of  $S$  called *equivalence classes*. If  $s \in S$ , then the *equivalence class* of  $s$  is the set of all elements related to it:

$$[s] = \{r \in S \mid r \sim s\}.$$

That is,  $[s]$  is the set of all elements that are considered “the same” as  $s$  under the relation  $\sim$ . For a given set  $S$  and an equivalence relation  $\sim$  on  $S$ , we define the *quotient set*, denoted  $S / \sim$ , whose elements are all the equivalence classes of elements in  $S$ . There is an obvious *quotient function* from  $S$  to  $S / \sim$  that maps each element  $s \in S$  to its equivalence class  $[s]$ .

**Definition 2.2.13.** Consider a pair  $(G, \Sigma)$ , where  $G$  is a class of ground types and  $\Sigma$  is a class of sorted operation symbols. A  $\lambda$ -theory is a triple  $((G, \Sigma), Ax)$ , where  $Ax$  is a class of equations-in-context over  $\lambda$ -terms constructed from  $(G, \Sigma)$ . The elements of  $Ax$  are called the *axioms* of the theory.

Let  $Th(Ax)$  denote the smallest class containing  $Ax$ , the equations presented in [Figure 2](#), and closed under exchange and substitution ([Lemma 2.2.8](#)). The elements of  $Th(Ax)$  are called the *theorems* of the theory.

For instance recall [Example 2.2.9](#):

$$- \triangleright (\lambda z : \mathbb{I} \otimes \mathbb{A}. \text{pm } z \text{ to } * \otimes y. y) (v \otimes w) : \mathbb{A} = w : \mathbb{A}$$

is a theorem.

## 2.2.9 Interlude: Booleans

Booleans can be represented as the type  $\mathbb{I} \oplus \mathbb{I}$ , in which case  $\text{True} = \text{inl}(*)$ ,  $\text{False} = \text{inr}(*)$  [\[31\]](#). We will use the equations in [Figure 2](#) to demonstrate that they possess certain properties that are characteristic of booleans in classical logic. The remaining properties can be proven through similar reasoning.

Given variables  $v : \mathbb{I} \oplus \mathbb{I}$  and  $w : \mathbb{I} \oplus \mathbb{I}$ , their conjunction and disjunction correspond to the following programs:

$$\mathbf{Conjunction} \triangleq v : \mathbb{I} \oplus \mathbb{I}, w : \mathbb{I} \oplus \mathbb{I} \triangleright \text{case } v \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . w; \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow \text{to } * . \text{dis}(w) \text{ to } * . \text{inr}_{\mathbb{I}}(*) \end{array} \right\}$$

$$\mathbf{Disjunction} \triangleq v : \mathbb{I} \oplus \mathbb{I}, w : \mathbb{I} \oplus \mathbb{I} \triangleright \text{case } v \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow \text{to } * . \text{dis}(w) \text{ to } * . \text{inl}_{\mathbb{I}}(*); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . w \end{array} \right\}$$

Moreover, negation can be expressed by the following program:

$$\mathbf{Negation} \triangleq v : \mathbb{I} \oplus \mathbb{I} \triangleright \text{case } v \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow \text{inr}_{\mathbb{I}}(x); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow \text{inl}_{\mathbb{I}}(y) \end{array} \right\}$$

**Lemma 2.2.14.**  $\text{inl}(*)$  acts as the neutral element for conjunction, whereas  $\text{inr}(*)$  serves as the absorbing element, i.e.

$$\mathbf{Conjunction} [\text{inl}(*)/v, w'/w] = w' \quad \text{and} \quad \mathbf{Conjunction} [\text{inr}(*)/v, w'/w] = \text{inr}(*)$$

*Proof.* These properties follow from the equations  $\beta_{case}^{inl}$ ,  $\beta_{case}^{inr}$ ,  $\beta_{\mathbb{I}_e}$ , and  $\eta_{dis}$ .

$$\begin{aligned}
& \mathbf{Conjunction} [\text{inl}(*)/v, w'/w] \\
&= \text{case } \text{inl}(*)\{ \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . w'; \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . \text{dis}(w') \text{ to } * . \text{inr}_{\mathbb{I}}(*) \} \\
&= * \text{ to } * . w' & (\beta_{case}^{inl}) \\
&= * \text{ to } * . w' & (\beta_{\mathbb{I}_e}) \\
&= w'
\end{aligned}$$

For the second equality to be well-defined, we observe that the context of  $w'$  must be empty. Moreover, we remark on the following:

$$\begin{aligned}
& x : \mathbb{I} \triangleright x \text{ to } * . \text{dis}(w) = x \text{ to } * . x : \mathbb{I} & (\eta_{dis}) \\
& \implies - \triangleright x \text{ to } * . \text{dis}(w)[*/x] = x \text{ to } * . x[*/x] : \mathbb{I} \\
& \implies - \triangleright * \text{ to } * . \text{dis}(w) = * \text{ to } * . * : \mathbb{I} \\
& \implies - \triangleright \text{dis}(w) = * : \mathbb{I} & (\beta_{\mathbb{I}_e})
\end{aligned}$$

With the equation above in hand, we have:

$$\begin{aligned}
& \mathbf{Conjunction} [\text{inr}(*)/v, w'/w] \\
&= \text{case } \text{inr}(*)\{ \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . w'; \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . \text{dis}(w') \text{ to } * . \text{inr}_{\mathbb{I}}(*) \} \\
&= * y \text{ to } * . \text{dis}(w') \text{ to } * . \text{inr}_{\mathbb{I}}(*) & (\beta_{case}^{inr}) \\
&= \text{dis}(w') \text{ to } * . \text{inr}_{\mathbb{I}}(*) & (\beta_{\mathbb{I}_e}) \\
&= * \text{ to } * . \text{inr}_{\mathbb{I}}(*) & (\eta_{dis} \text{ and } \beta_{\mathbb{I}_e}) \\
&= \text{inr}_{\mathbb{I}}(*)
\end{aligned}$$

Duvida na aplicação da regra do discard

□

Note that the idempotency property of conjunction (for  $\text{inl}(*)$  and  $\text{inr}(*)$ ) follows directly from the equalities above.

**Lemma 2.2.15.** *The conjunction of two terms is comutative, i.e.,*

$$\mathbf{Conjunction} [v'/v, w'/w] = \mathbf{Conjunction} [w'/v, v'/w]$$



*Proof.* This property follows from the equations  $\eta_{\text{case}}$ ,  $C_{\mathbb{I}_e}$ ,  $C_{\mathbb{I}_e}$ ,  $\eta_{\mathbb{I}_e}$ , and  $\eta_{\text{dis}}$ , and the  $\alpha$ -equivalence.

**Conjunction**  $[v'/v, w'/w]$

$$\begin{aligned}
&= \text{case } v \{ \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . w; \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . \text{dis}(w) \text{ to } * . \text{inr}_{\mathbb{I}}(*) \} [w'/w] \\
&= \text{case } w' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . \text{inl}_{\mathbb{I}}(a); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . \text{dis}(\text{inl}_{\mathbb{I}}(a)) \text{ to } * . \text{inr}_{\mathbb{I}}(*) \end{array} \right\} \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . \text{inr}_{\mathbb{I}}(b); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . \text{dis}(\text{inr}_{\mathbb{I}}(b)) \text{ to } * . \text{inr}_{\mathbb{I}}(*) \end{array} \right\} \end{array} \right\} \quad (\eta_{\text{case}}) \\
&= \text{case } w' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . \text{inl}_{\mathbb{I}}(a); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . a \text{ to } * . \text{inr}_{\mathbb{I}}(*) \end{array} \right\} \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . \text{inr}_{\mathbb{I}}(b); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . b \text{ to } * . \text{inr}_{\mathbb{I}}(*) \end{array} \right\} \end{array} \right\} \quad (\eta_{\text{dis}}) \\
&= \text{case } w' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . a \text{ to } * . \text{inl}_{\mathbb{I}}(*); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . a \text{ to } * . \text{inl}_{\mathbb{I}}(*) \end{array} \right\} \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . b \text{ to } * . \text{inr}_{\mathbb{I}}(*); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . b \text{ to } * . \text{inr}_{\mathbb{I}}(*) \end{array} \right\} \end{array} \right\} \quad (\eta_{\mathbb{I}_e}) \\
&= \text{case } w' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow a \text{ to } * . \text{inl}_{\mathbb{I}}(*) [x \text{ to } * . * / *]; \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow a \text{ to } * . \text{inl}_{\mathbb{I}}(*) [y \text{ to } * . * / *] \end{array} \right\} \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow b \text{ to } * . \text{inl}_{\mathbb{I}}(*) [x \text{ to } * . * / *]; \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow b \text{ to } * . \text{inl}_{\mathbb{I}}(*) [y \text{ to } * . * / *] \end{array} \right\} \end{array} \right\} \quad (C_{\mathbb{I}_e}) \\
&= \text{case } w' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow a \text{ to } * . \text{inl}_{\mathbb{I}}(x); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow a \text{ to } * . \text{inr}_{\mathbb{I}}(y) \end{array} \right\} \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow b \text{ to } * . x \text{ to } * . \text{inr}_{\mathbb{I}}(*); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow b \text{ to } * . y \text{ to } * . \text{inr}_{\mathbb{I}}(*) \end{array} \right\} \end{array} \right\} \quad (\eta_{\mathbb{I}_e}) \\
&= \text{case } w' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow a \text{ to } * . \text{inl}_{\mathbb{I}}(x); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow a \text{ to } * . \text{inr}_{\mathbb{I}}(y) \end{array} \right\} \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow b \text{ to } * . \text{dis}(\text{inr}_{\mathbb{I}}(x)) \text{ to } * . \text{inr}_{\mathbb{I}}(*); \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow b \text{ to } * . \text{dis}(\text{inr}_{\mathbb{I}}(y)) \text{ to } * . \text{inr}_{\mathbb{I}}(*) \end{array} \right\} \end{array} \right\} \quad (\eta_{\text{dis}})
\end{aligned}$$

$$\begin{aligned}
&= \text{case } w' \{ \text{inl}_{\mathbb{I}}(a) \Rightarrow a \text{ to } * . v'; \text{inr}_{\mathbb{I}}(b) \Rightarrow b \text{ to } * . \text{dis}(v') \text{ to } * . \text{inr}_{\mathbb{I}}(*) \} & (\eta_{\text{case}}) \\
&= \mathbf{Conjunction} [w'/v, v'/w]
\end{aligned}$$

□

To address the double negation property and De Morgan's laws, we begin by establishing a key equality known as the *syntactic fusion law*.

**Lemma 2.2.16** (*Syntactic fusion law*). *The following equality holds:*

$$v[(\text{case } a \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow w; \text{inr}_{\mathbb{A}}(y) \Rightarrow u \}) / z] = \text{case } a \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow v[w/z]; \text{inr}_{\mathbb{A}}(y) \Rightarrow v[u/z] \}.$$

*Proof.* This equality follows from the  $\alpha$ -equivalence and the equations  $\eta_{\text{case}}$ ,  $\beta_{\text{case}}^{\text{inl}}$ , and  $\beta_{\text{case}}^{\text{inr}}$ .

$$\begin{aligned}
&v[(\text{case } a \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow w; \text{inr}_{\mathbb{A}}(y) \Rightarrow u \}) / z] \\
&= v[(\text{case } a \{ \text{inl}_{\mathbb{B}}(x') \Rightarrow w[x'/x]; \text{inr}_{\mathbb{A}}(y') \Rightarrow u[y'/y] \}) / z] & (\alpha) \\
&= \text{case } a \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow v \left[ \text{case } \text{inl}_{\mathbb{I}}(x) \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x') \Rightarrow w[x'/x]; \\ \text{inr}_{\mathbb{I}}(y') \Rightarrow u[y'/y] \end{array} \right\} / z \right]; \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow v \left[ \text{case } \text{inr}_{\mathbb{I}}(y) \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x') \Rightarrow w[x'/x]; \\ \text{inr}_{\mathbb{I}}(y') \Rightarrow u[y'/y] \end{array} \right\} / z \right] \end{array} \right\} & (\eta_{\text{case}}) \\
&= \text{case } a \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow v[w[x'/x][x/x']/z]; \text{inr}_{\mathbb{A}}(y) \Rightarrow v[u[y'/y][y/y']/z] \} & (\beta_{\text{case}}^{\text{inl}} \text{ and } \beta_{\text{case}}^{\text{inr}}) \\
&= \text{case } a \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow v[w/z]; \text{inr}_{\mathbb{A}}(y) \Rightarrow v[u/z] \}
\end{aligned}$$

□

**Lemma 2.2.17.** *The double negation of a term  $w$  is equivalent to that term, i.e.,*

$$\mathbf{Negation}[\mathbf{Negation} [w/v]/v] = w.$$

*Proof.* This property follows from the syntactic fusion law and equations  $\beta_{\text{case}}^{\text{inl}}$ ,  $\beta_{\text{case}}^{\text{inr}}$ , and  $\eta_{\text{case}}$ .

$$\mathbf{Negation}[\mathbf{Negation} [w/v]/v]$$

$$\triangleq \mathbf{Negation}[\text{case } w \{ \text{inl}_{\mathbb{I}}(x) \Rightarrow \text{inr}_{\mathbb{I}}(x); \text{inr}_{\mathbb{I}}(y) \Rightarrow \text{inl}_{\mathbb{I}}(y) \} / v]$$

$$= \text{case } w \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow \mathbf{Negation}[\text{inr}_{\mathbb{I}}(x)/v]; \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow \mathbf{Negation}[\text{inl}_{\mathbb{I}}(y)/v] \end{array} \right\} \quad (\text{syntactic fusion law})$$

$$= \text{case } w \{ \text{inl}_{\mathbb{I}}(x) \Rightarrow \text{inr}_{\mathbb{I}}(x); \text{inl}_{\mathbb{I}}(x) \Rightarrow \text{inr}_{\mathbb{I}}(y) \} \quad (\beta_{case}^{inl} \text{ and } \beta_{case}^{inr})$$

$$= w \quad (\eta_{case})$$

□

**Lemma 2.2.18.** *De Morgan's laws hold for terms  $\Gamma \triangleright v' : \mathbb{I} \oplus \mathbb{I}$  and  $\Delta \triangleright w : \mathbb{I} \oplus \mathbb{I}$ , i.e.,*

$$\mathbf{Disjunction}[\mathbf{Negation} [v'/v]/v, \mathbf{Negation} [w'/v]/w] = \mathbf{Negation}[\mathbf{Conjunction}[v'/v, w'/w]/v]$$

and

$$\mathbf{Conjunction}[\mathbf{Negation} [v'/v]/v, \mathbf{Negation} [w'/v]/w] = \mathbf{Negation}[\mathbf{Disjunction}[v'/v, w'/w]/v].$$

*Proof.* Both De Morgan's laws follow from the equations  $\eta_{case}, \beta_{case}^{inl}, \beta_{case}^{inr}, c_{\mathbb{I}_e}, \eta_{dis}$ , the syntatic fusion law (SFL), and the  $\alpha$ -equivalence.

$$\mathbf{Disjunction}[\mathbf{Negation} [v'/v]/v, \mathbf{Negation} [w'/v]/w]$$

$$\triangleq \text{case case } v \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{inr}_{\mathbb{I}}(a); \\ \text{inl}_{\mathbb{I}}(b) \Rightarrow \text{inl}_{\mathbb{I}}(b) \end{array} \right\} \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow \dots; \\ \text{inl}_{\mathbb{I}}(y) \Rightarrow \dots \end{array} \right\} [v/v']$$

$$= \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(c) \Rightarrow \text{case case inl}_{\mathbb{I}}(c) \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{inr}_{\mathbb{I}}(a); \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{inl}_{\mathbb{I}}(b) \end{array} \right\} \dots; \\ \text{inl}_{\mathbb{I}}(d) \Rightarrow \text{case case inr}_{\mathbb{I}}(d) \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{inr}_{\mathbb{I}}(a); \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{inl}_{\mathbb{I}}(b) \end{array} \right\} \dots \end{array} \right\} \quad (\eta_{case})$$

$$= \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(c) \Rightarrow \text{case inr}_{\mathbb{I}}(c) \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow \dots; \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow \dots \end{array} \right\}; \\ \text{inl}_{\mathbb{I}}(d) \Rightarrow \text{case inl}_{\mathbb{I}}(d) \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(x) \Rightarrow \dots; \\ \text{inr}_{\mathbb{I}}(y) \Rightarrow \dots \end{array} \right\} \end{array} \right\} \quad (\beta_{case}^{inl} \text{ and } \beta_{case}^{inr})$$

$$\begin{aligned}
&= \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(c) \Rightarrow c \text{ to } * . \mathbf{Negation}[w'/v]; \\ \text{inl}_{\mathbb{I}}(d) \Rightarrow d \text{ to } * . \text{dis} \left( \text{case } w' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{inr}_{\mathbb{I}}(a); \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{inl}_{\mathbb{I}}(b) \end{array} \right\} \right) \dots \end{array} \right\} & (\beta_{case}^{inl}) \\
& & \text{and} \\
& & (\beta_{case}^{inr}) \\
&= \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(c) \Rightarrow c \text{ to } * . \mathbf{Negation}[w'/v]; \\ \text{inl}_{\mathbb{I}}(d) \Rightarrow d \text{ to } * . \text{dis}(w') \text{ to } * . \text{inl}_{\mathbb{I}}(*) \end{array} \right\} & (\eta_{dis}) \\
&= \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(c) \Rightarrow c \text{ to } * . \mathbf{Negation}[w'/v]; \\ \text{inl}_{\mathbb{I}}(d) \Rightarrow d \text{ to } * . \text{dis}(w') \text{ to } * . \text{inr}_{\mathbb{I}}(*) \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(a) \Rightarrow \text{inr}_{\mathbb{I}}(a); \\ \text{inr}_{\mathbb{I}}(b) \Rightarrow \text{inl}_{\mathbb{I}}(b) \end{array} \right\} \end{array} \right\} & (\beta_{case}^{inr}) \\
&= \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(c) \Rightarrow c \text{ to } * . \mathbf{Negation}[w'/v]; \\ \text{inl}_{\mathbb{I}}(d) \Rightarrow d \text{ to } * . \text{dis}(w') \text{ to } * . \mathbf{Negation}[\text{inr}_{\mathbb{I}}(*)/v] \end{array} \right\} \\
&= \text{case } v' \left\{ \begin{array}{l} \text{inl}_{\mathbb{I}}(c) \Rightarrow \mathbf{Negation}[c \text{ to } * . w'/v]; \\ \text{inl}_{\mathbb{I}}(d) \Rightarrow \mathbf{Negation}[d \text{ to } * . \text{dis}(w') \text{ to } * . \text{inr}_{\mathbb{I}}(*)/v] \end{array} \right\} & (c_{\mathbb{I}_e}) \\
&= \mathbf{Negation} [\text{case } v' \{ \text{inl}_{\mathbb{I}}(c) \Rightarrow c \text{ to } * . w; \text{inr}_{\mathbb{I}}(d) \Rightarrow d \text{ to } * . \text{dis}(w) \text{ to } * . \text{inr}_{\mathbb{I}}(*) \} / v] & (\text{SFL}) \\
&= \mathbf{Negation} [\mathbf{Conjunction}[v'/v, w', w] / v] & (\alpha)
\end{aligned}$$

□

### 2.2.10 Metric equational system

*Metric equations* [38, 39] are a strong candidate for reasoning about approximate program equivalence. These equations take the form of  $t =_{\epsilon} s$ , where  $\epsilon$  is a non-negative rational representing the “maximum distance” between the two terms  $t$  and  $s$ . The metric equational system for affine lambda calculus is depicted in Figure 3. Strictly speaking, the equations  $\Gamma \triangleright v = w : A$  in Figure 2, which in this setting abbreviate  $\Gamma \triangleright w =_0 v : A$ , are also part of the metric equational system.

$$\begin{array}{c}
\frac{}{v =_0 v} \text{ (refl)} \qquad \frac{v =_q w \quad w =_r u}{v =_{q+r} u} \text{ (trans)} \qquad \frac{v =_q w \quad r \geq q}{v =_r w} \text{ (weak)} \\
\\
\frac{\forall r > q. v =_r w}{v =_q w} \text{ (arch)} \qquad \frac{\forall i \leq n. v =_{q_i} w}{v =_{\wedge q_i} w} \text{ (join)} \qquad \frac{v =_q w}{w =_q v} \text{ (sym)} \\
\\
\frac{v =_q w \quad v' =_r w'}{v \otimes v' =_{q+r} w \otimes w'} \qquad \frac{\forall i \leq n. v_i =_{q_i} w_i}{f(v_1, \dots, v_n) =_{\Sigma q_i} f(w_1, \dots, w_n)} \qquad \frac{v =_q w}{\lambda x : \mathbb{A}. v =_q \lambda x : \mathbb{A}. w} \\
\\
\frac{v =_q w \quad v' =_r w'}{\text{pm } v \text{ to } x \otimes y. v' =_{q+r} \text{pm } w \text{ to } x \otimes y. w'} \quad \frac{v =_q w}{\text{dis}(v) =_q \text{dis}(w)} \quad \frac{v =_q w \quad v' =_r w'}{v v' =_{q+r} w w'} \\
\\
\frac{\Gamma \triangleright v =_q w : \mathbb{A} \quad \Delta \in \text{perm}(\Gamma)}{\Delta \triangleright v =_q w : \mathbb{A}} \quad \frac{v =_q w \quad v' =_r w'}{v \text{ to } * . v' =_{q+r} w \text{ to } * . w'} \quad \frac{v =_q w \quad v' =_r w'}{v[v'/x] =_{q+r} w[w'/x]}
\end{array}$$

Figure 3: Metric equational system

Here,  $\text{perm}(\Gamma)$  denotes the set of possible permutations of context  $\Gamma$ . The rules (refl), (trans), and (sym) generalize the properties of reflexivity, transitivity, and symmetry of equality. Rule (weak) asserts that if two terms are at a maximum distance  $q$  from each other, then they are also separated by any  $r \geq q$ . Rule (arch) states that if  $v =_r w$  for all approximations  $r$  of  $q$ , then it necessarily follows that  $v =_q w$ . The rule (join) expresses that if several maximum distances between two terms are known, the actual maximum distance between them is the minimum of these distances. The rule that follows conveys that if the maximum distance between two terms  $v$  and  $w$  is  $q$ , and the maximum distance between terms  $v'$  and  $w'$  is  $r$ , then the maximum distance between the tensor products  $v \otimes v'$  and  $w \otimes w'$  is  $q + r$ . The remaining rules follow similar reasoning.

Não sei se dá para postular para  $f$  sem argumento, não dá para usar nenhuma das regras, a da substituição não dá para aplicar porque  $f$  sozinho não é termo

**Example 2.2.19.** To illustrate the usefulness of these equations, consider the program  $P$  that receives a tensor product, swaps its elements and then applies a function  $f : \mathbb{A} \rightarrow \mathbb{D} \in \Sigma$  to the new second element of the tensor pair:

$$P = x : \mathbb{A}, y : \mathbb{B} \triangleright \text{pm } x \otimes y \text{ to } a \otimes b. b \otimes f(a) : \mathbb{B} \otimes \mathbb{D}$$

Let  $f^\epsilon$  be an erroneous implementation of  $f$ . The program above is thus rewritten as:

$$P^\epsilon = x : \mathbb{A}, y : \mathbb{B} \triangleright \text{pm } x \otimes y \text{ to } a \otimes b. b \otimes f(a)^\epsilon : \mathbb{B} \otimes \mathbb{D}$$

Consider we can postulate the axiom  $f^\epsilon(a) =_\epsilon f(a)$ . Then, it is possible to show that  $P^\epsilon =_\epsilon P$  using our metric equational system. The prove is as follows. The types and contexts are omitted for brevity as no ambiguity arises.

$$\begin{array}{ll} 1 & f(a)^\epsilon =_\epsilon f(a) \\ 2 & b =_0 b \quad (\text{refl}) \\ 3 & b \otimes f(a)^\epsilon =_\epsilon b \otimes f(a) \quad (1, 2, \otimes_i) \\ 4 & x \otimes y =_0 x \otimes y \quad (\text{refl}) \\ 5 & \text{pm } x \otimes y \text{ to } a \otimes b. b \otimes f(a)^\epsilon =_\epsilon \text{pm } x \otimes y \text{ to } a \otimes b. b \otimes f(a) \quad (3, 4, \otimes_e) \end{array}$$

Note that without a metric equation for conditionals, we cannot, for instance, reason about approximate equivalence in programs such as **Conjunction**, discussed in the previous subsection. For example, it would be interesting to know whether, given that  $v' =_\epsilon \text{inl}_{\mathbb{I}}(*)$ , it follows that

$$\mathbf{Conjunction}[v/v', w/w'] =_\epsilon \mathbf{Conjunction}[\text{inl}_{\mathbb{I}}(*)/v', w/w'].$$

**Definition 2.2.20.** Consider a tuple  $(G, \Sigma)$ , where  $G$  is a class of ground types and  $\Sigma$  is a class of sorted operation symbols of the form  $f : A_1, \dots, A_n \rightarrow A$  with  $n \geq 1$ . A *metric  $\lambda$ -theory* is a tuple  $((G, \Sigma), Ax)$ , where  $Ax$  is a class of *metric equations-in-context* over  $\lambda$ -terms constructed from  $(G, \Sigma)$ .

The elements of  $Ax$  are called the *axioms* of the theory. Let  $Th(Ax)$  denote the smallest class that contains  $Ax$  and is closed under the rules presented in [Figure 2](#) (i.e., the classical equational system) and [Figure 3](#). The elements of  $Th(Ax)$  are called the *theorems* of the theory.

For instance, in [Example 2.2.19](#),  $\mathbf{Conjunction}[v/v', w/w'] =_\epsilon \mathbf{Conjunction}[\text{inl}_{\mathbb{I}}(*)/v', w/w']$  is a theorem.

## 2.3 Category theory

Cenas para a descrição da secção: ref `awodeyCategoryTheory2010` e nota sobre a maioria dos exemplos serem retirados do `noson`

Category theory originated as an effort to connect and unify two distinct areas of mathematics. The goal was to study and classify specific geometric structures—such as topological spaces, manifolds, and bundles—by associating them with corresponding algebraic structures like groups, rings, and abelian groups. It became clear that a language was needed to connect geometric and algebraic objects—one not explicitly tailored to geometry or algebra. Only a language of such generality could allow meaningful discussion across both fields. This is the birth of category theory. Described as “a language about nothing, and therefore about everything,” category theory provides a highly general way of discussing mathematical concepts. It was invented by Samuel Eilenberg and Saunders MacLane [40]. They organized various mathematical structures into categories called geometric others algebraic. To connect these categories, they defined functors, which map objects and morphisms from one category to another, much like functions do. They further introduced natural transformations, which provide a way to compare functors, translating the results of one functor into those of another. [32]

### 2.3.1 Categories

**Definition 2.3.1.** A *category*  $C$  consists of

- a collection of objects  $A, B, C, \dots$ , denoted  $|C|$  or  $\text{Obj}(C)$ ;
- a collection of morphisms  $f, g, \dots$ , usually denoted  $C(A, B)$ ,  $\text{Hom}_C(A, B)$ , or  $\text{Hom}(A, B)$  if there is no ambiguity.

The collection for morphisms has the following structure:

- Each morphism has a specified domain and codomain; the notation  $f : A \rightarrow B$  indicates that  $f$  is a morphism from object  $A$  to object  $B$ .
- Every object  $A$  has an identity morphism  $\text{id}_A : A \rightarrow A$ .
- For any pair of morphisms  $f : A \rightarrow B$  and  $g : B \rightarrow C$ , where the codomain of  $f$  matches the domain of  $g$ , there exists a composite morphism  $g \circ f : A \rightarrow C$ . We will also write  $g \circ f$  as  $f \cdot g$  or simply  $fg$ .

The composition is required to satisfy the two following laws: if  $f : A \rightarrow B$ ,  $g : B \rightarrow C$ , and  $h : C \rightarrow D$  are morphisms, then

- $f \circ \text{id}_a = f = \text{id}_b \circ f$ ;
- $(f \circ g) \circ h = f \circ (g \circ h)$ .

**Example 2.3.2.** Set is the category whose objects are sets and whose morphisms are functions between them. Given a function  $f : A \rightarrow B$ , it assigns to each element  $a \in A$  a unique image  $f(a) \in B$ . For any two functions  $f : A \rightarrow B$  and  $g : B \rightarrow C$ , their composition is defined by

$$(g \circ f)(a) = g(f(a)) \quad \text{for all } a \in A.$$

This composition is *associative*. That is, for any further function  $h : C \rightarrow D$ , we have

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

since for every  $a \in A$ ,

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

Moreover, for every set  $A$ , there exists an *identity function*

$$\text{id}_A : A \rightarrow A, \quad \text{defined by } \text{id}_A(a) = a,$$

which satisfies the unit laws for composition:

$$f \circ \text{id}_A = f \quad \text{and} \quad \text{id}_B \circ f = f$$

for any function  $f : A \rightarrow B$ .

Therefore, **Set**, with sets as objects and functions as morphisms, satisfies the axioms of a category.

Another common type of example consists of categories of sets equipped with additional structure, along with functions that preserve that structure.

**Definition 2.3.3.** A *partially ordered set* or *partial order* is a set  $A$  equipped with a binary relation  $\leq_A$  satisfying the following properties for all  $a, b, c \in A$ :

- Reflexivity:  $a \leq_A a$ ;
- Transitivity: If  $a \leq_A b$  and  $b \leq_A c$ , then  $a \leq_A c$ ;
- Antisymmetry: If  $a \leq_A b$  and  $b \leq_A a$ , then  $a = b$ .



**Example 2.3.4.** The set of real numbers  $\mathbb{R}$ , equipped with the usual ordering  $\leq$ , forms a poset. Moreover, it is *linearly ordered* (or *totally ordered*), since for any  $x, y \in \mathbb{R}$ , either  $x \leq y$  or  $y \leq x$  holds.

**Definition 2.3.5.** Given two partial orders  $(A, \leq_A)$  and  $(B, \leq_B)$ , a function  $m : A \rightarrow B$  is called a *monotone map* (or *order-preserving map*) if for all  $a, a' \in A$ ,

$$a \leq_A a' \Rightarrow m(a) \leq_B m(a').$$

**Example 2.3.6.** PO is the category of all partial orders and all monotone maps. First, for any poset  $A$ , the identity function  $\text{id}_A : A \rightarrow A$  is monotone. Indeed, for all  $a \in A$ ,

$$a \leq_A a \Rightarrow \text{id}_A(a) \leq_A \text{id}_A(a).$$

Next, given monotone maps  $f : A \rightarrow B$  and  $g : B \rightarrow C$ , their composition  $g \circ f : A \rightarrow C$  is also monotone. For all  $a, a' \in A$ , if  $a \leq_A a'$ , then

$$f(a) \leq_B f(a') \quad \text{and} \quad g(f(a)) \leq_C g(f(a')),$$

so it follows that

$$(g \circ f)(a) \leq_C (g \circ f)(a').$$

**Example 2.3.7.** Each partially ordered set naturally defines a category. Let  $(P, \leq)$  be a poset. We define a category  $B(P, \leq)$ , often denoted simply by  $B(P)$  or even  $P$ , where the objects are the elements of  $P$ , and there is a unique morphism  $p \rightarrow q$  if and only if  $p \leq q$ . The reflexivity of the order  $\leq$  ensures the existence of identity morphisms, while transitivity guarantees that morphisms compose appropriately. Moreover, since there is at most one morphism between any two objects, composition is trivially associative.

**Example 2.3.8.** CVect is the category of finite complex vector spaces and linear mappings.

**Example 2.3.9.** Given a functional programming language  $L$ , we can define a category CompFunc.

In this category, the objects represent the data types of the language  $L$ , and the morphisms correspond to computable functions or programs. A function is considered *computable* if a computer program is capable of executing that function.

The composition of morphisms is defined as follows: given two morphisms  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$ , the composition  $g \circ f : X \rightarrow Z$  is defined by applying  $g$  to the output of  $f$ . This composition is often written as  $f; g$ .

Additionally, the identity morphism  $\text{id}_X : X \rightarrow X$  represents the “identity program,” which returns its input without making any changes (i.e., it “does nothing”).

**Definition 2.3.10.** A morphism  $f : A \rightarrow B$  in a category  $\mathcal{C}$  is called an *isomorphism* if there exists a morphism  $f^{-1} : B \rightarrow A$  such that

$$f^{-1} \circ f = \text{id}_A \quad \text{and} \quad f \circ f^{-1} = \text{id}_B.$$

In this case,  $f^{-1}$  is called the *inverse* of  $f$ , and it is unique. If such an isomorphism exists, we say that  $A$  and  $B$  are *isomorphic*, written  $A \cong B$ .

One of the central ideas in category theory is *duality*. Simply put, for a given definition of a structure, there is often a corresponding dual concept obtained by reversing the directions of all the arrows.

**Definition 2.3.11.** Let  $\mathcal{C}$  be a category. The *opposite category*, denoted  $\mathcal{C}^{\text{op}}$ , is defined as follows:

- The objects of  $\mathcal{C}^{\text{op}}$  are the same as those of  $\mathcal{C}$ .
- For any pair of objects  $A, B$ , the hom-set in  $\mathcal{C}^{\text{op}}$  is defined by

$$\text{Hom}_{\mathcal{C}^{\text{op}}}(A, B) = \text{Hom}_{\mathcal{C}}(B, A),$$

that is, each morphism  $f : A \rightarrow B$  in  $\mathcal{C}^{\text{op}}$  corresponds to a morphism  $f : B \rightarrow A$  in  $\mathcal{C}$ .

- Composition in  $\mathcal{C}^{\text{op}}$  is defined using the composition in  $\mathcal{C}$ , but in reverse order. That is, if

$$A \xrightarrow{f} B \xrightarrow{g} C$$

are morphisms in  $\mathcal{C}^{\text{op}}$ , corresponding to morphisms

$$C \xrightarrow{g} B \xrightarrow{f} A$$

in  $\mathcal{C}$ , then the composition in  $\mathcal{C}^{\text{op}}$  is defined by

$$g \circ f := f \circ_{\mathcal{C}} g.$$

Thus,  $\mathcal{C}^{\text{op}}$  reverses the direction of morphisms and composition while retaining the same collection of objects.

**Definition 2.3.12.** A subcategory  $D$  of a category  $C$  is a category such that

- All the objects of  $D$  are objects of  $C$ , and all the morphisms of  $D$  are morphisms of  $C$  (that is,  $D_0 \subseteq C_0$  and  $D_1 \subseteq C_1$ ).
- The domain and codomain of any morphism in  $D$  are the same as in  $C$  (in other words, the domain and codomain maps for  $D$  are the restrictions of those for  $C$ ).
- It follows that for any objects  $A$  and  $B$  in  $D$ , we have  $\text{Hom}_D(A, B) \subseteq \text{Hom}_C(A, B)$ . If  $A$  is an object of  $D$ , then its identity morphism  $\text{id}_A$  in  $C$  also belongs to  $D$ .
- If  $f : A \rightarrow B$  and  $g : B \rightarrow C$  are morphisms in  $D$ , then the composite  $g \circ f$ , as defined in  $C$ , is also in  $D$ , and coincides with the composite in  $D$ .

**Example 2.3.13.** The category  $\text{FinSet}$ , whose objects are finite sets and whose morphisms are functions between them, forms a subcategory of the category  $\text{Set}$ .

**Definition 2.3.14.** A category is called *small* if both its collection of objects and its collection of morphisms form sets. A category is called *locally small* if, for every pair of objects, the corresponding hom-set is a set.

### 2.3.2 Products and coproducts

A category frequently possesses more intricate structure than a mere collection of objects and their morphisms. The existence of particular relationships among certain objects and morphisms can make some objects have important properties.

It should be noted that a diagram is said to commute if, for every pair of objects  $X$  and  $Y$  in the diagram, all directed paths from  $X$  to  $Y$  yield equal morphisms.

**Definition 2.3.15.** An object  $0$  in a category  $C$  is called an *initial object* if for every object  $A \in C$ , there exists a unique morphism  $f : 0 \rightarrow A$ .

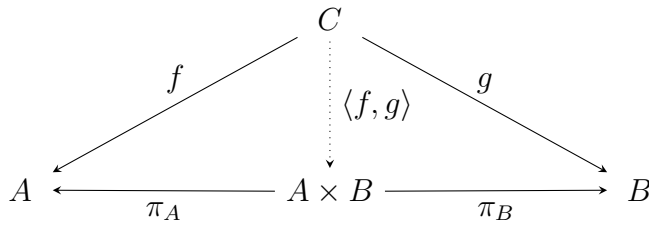
**Definition 2.3.16.** An object  $T$  in a category  $C$  is called a *terminal object* if for every object  $A \in C$ , there exists a unique morphism  $f : A \rightarrow T$ .

**Example 2.3.17.** In the category  $\text{Set}$ , the empty set  $\emptyset$  is an initial object, since for any set  $S$ , there exists a unique function  $f : \emptyset \rightarrow S$ . This function is unique because there are no elements in  $\emptyset$  to map.

Any singleton set, such as  $\{*\}$  or  $\{a\}$ , is a terminal object in this category. For any set  $S$ , there exists a unique function  $f : S \rightarrow \{*\}$ , which maps every element of  $S$  to the sole element of the singleton set

**Example 2.3.18.** Let  $(P, \leq)$  be a partial order and  $P$  be its associated category. Here, the initial object is the *bottom element*—an element that is less than or equal to every other element in  $P$ . The terminal object in  $P$  is the *top element*—an element that is greater than or equal to every other element in  $P$ .

**Definition 2.3.19 (Product).** Consider a category  $C$ . We say that it has (binary) products if for any objects  $A$  and  $B$  in  $C$  there also exists an object  $A \times B$  in  $C$  with morphisms  $\pi_A : A \times B \rightarrow A$  and  $\pi_B : A \times B \rightarrow B$  that satisfy a certain universal property: specifically for every two morphisms  $f : C \rightarrow A$  and  $g : C \rightarrow B$  there exists a *unique* morphism  $\langle f, g \rangle : C \rightarrow A \times B$  called *pairing* that makes the diagram below commute.

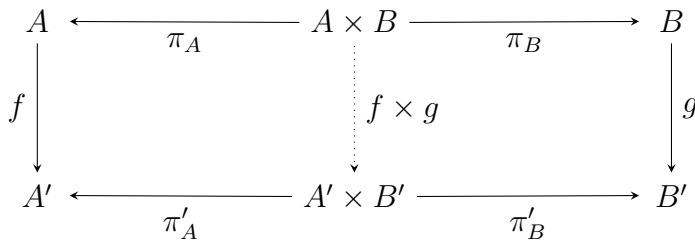


**Definition 2.3.20.** Let  $A \times B$  be a product of objects  $A$  and  $B$ , and let  $A' \times B'$  be a product of objects  $A'$  and  $B'$  in a category  $C$ . Suppose we are given morphisms  $f : A \rightarrow A'$  and  $g : B \rightarrow B'$ .

Then there exists a unique morphism

$$f \times g : A \times B \rightarrow A' \times B'$$

such that the following diagram commutes.



This induced morphism  $f \times g$  is called the *product of the morphisms  $f$  and  $g$* , and it is given explicitly by

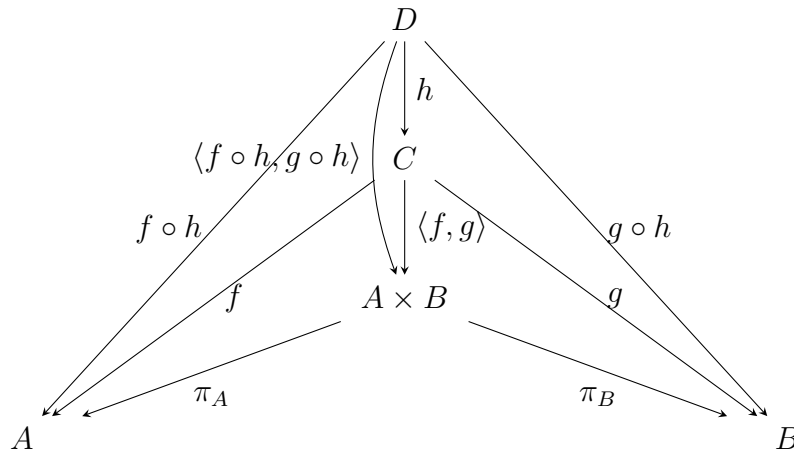
$$f \times g = \langle f \circ \pi_A, g \circ \pi_B \rangle.$$

**Theorem 2.3.21.** Let  $A \times B$  be the product of objects  $A$  and  $B$  in a category  $\mathcal{C}$ . For any object  $C$  and morphisms  $f : C \rightarrow A$  and  $g : C \rightarrow B$  are morphisms, it holds that:

$$\langle f \circ h, g \circ h \rangle = \langle f, g \rangle \circ h.$$

*Proof.* The universal property of the product induces a unique morphism  $\langle f, g \rangle : C \rightarrow A \times B$  such that  $\pi_A \circ \langle f, g \rangle = f$  and  $\pi_B \circ \langle f, g \rangle = g$ .

Now, let  $h : D \rightarrow C$  be another morphism. Then the compositions  $f \circ h : D \rightarrow A$  and  $g \circ h : D \rightarrow B$  also induce a unique morphism  $\langle f \circ h, g \circ h \rangle : D \rightarrow A \times B$  by the universal property of the product. As a result, by the universal property of the product the following diagram commutes.



□

**Example 2.3.22.** In the category  $\mathbf{Set}$ , the product of two sets  $A$  and  $B$  is given by their Cartesian product, denoted

$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$

The projection maps are defined by

$$\pi_A(a, b) = a \quad \text{and} \quad \pi_B(a, b) = b.$$

Given a set  $C$  and morphisms  $f : C \rightarrow A$  and  $g : C \rightarrow B$ , their pairing is the map

$$\langle f, g \rangle(c) = (f(c), g(c)).$$

**Example 2.3.23.** Let  $(P, \leq)$  be a partial order and  $P$  be its associated category. Consider a product of elements  $p \times q \in P$ . Then, by definition, there must exist projections satisfying

$$p \times q \leq p \quad \text{and} \quad p \times q \leq q.$$

Furthermore, for any element  $x \in P$ , if

$$x \leq p \quad \text{and} \quad x \leq q,$$

then it follows that

$$x \leq p \times q.$$

This operation  $p \times q$  corresponds to what is commonly known as the *greatest lower bound* or *meet*, and is typically denoted by  $p \wedge q$ .

Maybe tirar este exemplo

**Example 2.3.24.** In the category  $\mathbf{CVect}$ , the product of two vector spaces  $V$  and  $W$  corresponds to their direct sum, denoted by  $V \oplus W$ . The projection maps are the linear maps

$$\pi_V : V \oplus W \rightarrow V, \quad \pi_V(v, w) = v,$$

$$\pi_W : V \oplus W \rightarrow W, \quad \pi_W(v, w) = w.$$

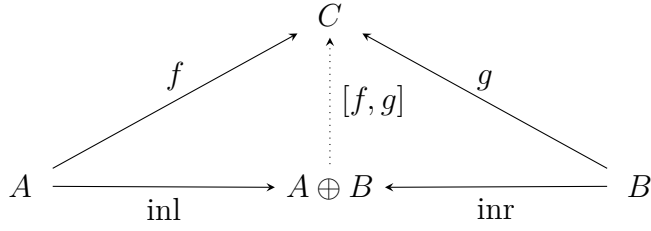
Given any vector space  $U$  and linear maps  $f : U \rightarrow V$  and  $g : U \rightarrow W$ , the unique map  $\langle f, g \rangle : U \rightarrow V \oplus W$  is defined by

$$\langle f, g \rangle(u) = (f(u), g(u)).$$

The *coproduct* is the dual of the *product*—it is obtained by reversing all the morphisms in the definition of a product. Consequently, a product in a category  $\mathbf{C}$  corresponds to a coproduct in the opposite category  $\mathbf{C}^{\text{op}}$ .

**Definition 2.3.25.** Consider a category  $\mathbf{C}$ . We say that it has (binary) coproducts if for any objects  $A$  and  $B$  in  $\mathbf{C}$  there also exists an object  $A \oplus B$  in  $\mathbf{C}$  with morphisms  $\text{inl} : A \rightarrow A \oplus B$  and  $\text{inr} : B \rightarrow A \oplus B$  that satisfy a certain universal property: specifically for every two morphisms  $f : A \rightarrow C$  and  $g : B \rightarrow C$  there exists a *unique* morphism  $[f, g] : A \oplus B \rightarrow C$

known as *co-pairing* that makes the diagram below commute.

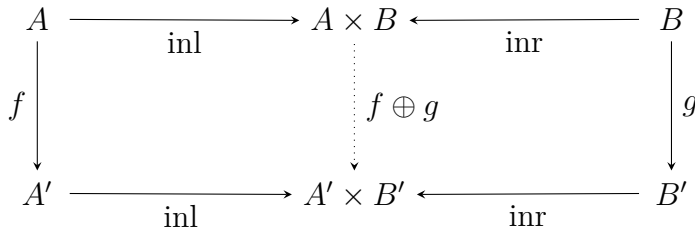


**Definition 2.3.26.** Let  $A \oplus B$  be a coproduct of objects  $A$  and  $B$ , and let  $A' \times B'$  be a co-product of objects  $A'$  and  $B'$  in a category  $\mathcal{C}$ . Suppose we are given morphisms  $f : A \rightarrow A'$  and  $g : B \rightarrow B'$ .

Then there exists a unique morphism

$$f \times g : a \times b \rightarrow a' \times b'$$

such that the following diagram commutes.



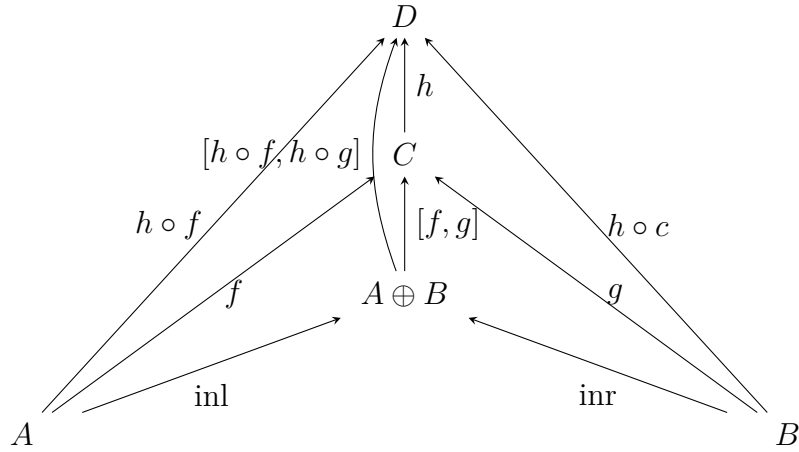
This induced morphism  $f \oplus g$  is called the *coproduct of the morphisms  $f$  and  $g$* , and it is given explicitly by

$$f \oplus g = [\text{inl} \circ f, \text{inr} \circ g].$$

**Theorem 2.3.27.** Let  $A \oplus B$  be the product of objects  $A$  and  $B$  in a category  $\mathcal{C}$ . For any object  $C$  and morphisms  $f : A \rightarrow C$  and  $g : B \rightarrow C$  are morphisms, it holds that:

$$[h \circ f, h \circ g] = h \circ [f, g].$$

*Proof.* By the universal property of the coprodut the following diagram commutes.



□

**Example 2.3.28.** In the category  $\text{Set}$ , the coproduct  $A \oplus B$  of two sets is their disjoint union, which can be constructed as

$$A \oplus B = \{(a, 1) \mid a \in A\} \cup \{(b, 2) \mid b \in B\}.$$

The canonical coproduct injections are defined by

$$i_1(a) = (a, 1), \quad i_2(b) = (b, 2).$$

Given any set  $C$  and functions  $f : A \rightarrow C$  and  $g : B \rightarrow C$ , the copairing  $[f, g] : A \oplus B \rightarrow C$  is defined by

$$[f, g](x, \delta) = \begin{cases} f(x) & \text{if } \delta = 1, \\ g(x) & \text{if } \delta = 2. \end{cases}$$

**Example 2.3.29.** Let  $(P, \leq)$  be a partial order and  $\mathbf{P}$  be its associated category. Consider a coproduct of elements  $p \oplus q \in P$ . Then, by definition, there must exist injections satisfying

$$p \leq p + q \quad \text{and} \quad q \leq p + q.$$

Furthermore, for any element  $z \in P$ , if

$$p \leq z \quad \text{and} \quad q \leq z,$$

then it follows that

$$p + q \leq z.$$

This operation  $p + q$  corresponds to what is commonly known as the *least upper bound* or *join*, and is typically denoted by  $p \vee q$ .



Maybe tirar este exemplo

**Example 2.3.30.** In  $\mathbf{CVect}$ , the coproduct coincides with the product. In such cases, this structure is called a *biproduct*. In both  $\mathbf{CVect}$ , the injection maps are the linear maps

$$\text{inl} : V \rightarrow V \oplus W, \quad \text{inl}(v) = (v, 0),$$

$$\text{inr} : W \rightarrow V \oplus W, \quad \text{inr}(w) = (0, w).$$

Given any vector space  $U$  and linear maps  $f : V \rightarrow U$  and  $g : W \rightarrow U$ , the unique map  $\langle f, g \rangle : V \oplus W \rightarrow U$  is defined by

$$[f, g](v, w) = f(v) + g(w).$$

### 2.3.3 Functors

Although categories are interesting on their own, the real strength of category theory lies in understanding how categories relate to one another. Just as functions express relationships between sets, functors play a similar role for categories. A functor maps each object in one category to an object in another category, and it does the same for morphisms, preserving the structure of the relationships.

**Definition 2.3.31.** Let  $\mathbf{C}$  and  $\mathbf{D}$  be two categories. A *functor*  $F : \mathbf{C} \rightarrow \mathbf{D}$  consists of a mapping that assigns to each object  $A$  in  $\mathbf{C}$  an object  $FA$  in  $\mathbf{D}$ , and to each morphism  $f \in \text{Hom}_{\mathbf{C}}(A, B)$  a morphism  $Ff \in \text{Hom}_{\mathbf{D}}(FA, FB)$ , in such a way that the following two conditions are satisfied for all objects  $A, B, C$  in  $\mathbf{C}$  and all morphisms  $f \in \text{Hom}_{\mathbf{C}}(A, B)$  and  $g \in \text{Hom}_{\mathbf{C}}(B, C)$ :

$$F(\text{id}_A) = \text{id}_{FA}, \quad F(g \circ f) = F(g) \circ F(f).$$

A functor  $F : \mathbf{C} \rightarrow \mathbf{D}$  is said to be *full* if, for all objects  $A$  and  $B$  in  $\mathbf{C}$ , the induced map

$$F_{A,B} : \text{Hom}_{\mathbf{C}}(A, B) \longrightarrow \text{Hom}_{\mathbf{D}}(FA, FB), \quad f \mapsto Ff,$$

is surjective. The functor is called *faithful* if each  $F_{A,B}$  is injective, and *fully faithful* if each  $F_{A,B}$  is bijective.

A *full embedding* is a functor that is fully faithful and, in addition, injective on objects.

**Example 2.3.32.** Let  $\mathbf{C}$  be a category. Then there exists an *identity functor*  $\text{id}_{\mathbf{C}} : \mathbf{C} \rightarrow \mathbf{C}$ , which is defined on objects by  $\text{id}_{\mathbf{C}}(A) = A$  for every object  $A$  in  $\mathbf{C}$ , and analogously on morphisms, that is,  $\text{id}_{\mathbf{C}}(f) = f$  for every morphism  $f$  in  $\mathbf{C}$ .

**Example 2.3.33.** Consider the natural numbers  $\mathbb{N}$  as partial order category. There is a functor  $(-) + 5 : \mathbb{N} \rightarrow \mathbb{N}$  that maps each object  $m \in \mathbb{N}$  to  $m + 5$ . This defines a functor because it preserves morphisms: if  $m \leq m'$ , then  $m + 5 \leq m' + 5$ . Moreover, the identity morphisms are preserved, since  $(\text{id}_m) + 5 = \text{id}_{m+5}$ .

**Example 2.3.34.** Consider the set of real numbers  $\mathbb{R}$  and the set of integers  $\mathbb{Z}$ , each regarded as a partial order category. In this context, there exists a functor  $\text{Floor} : \mathbb{R} \rightarrow \mathbb{Z}$  that assigns to each real number  $r \in \mathbb{R}$  the greatest integer less than or equal to  $r$ , denoted  $\lfloor r \rfloor$ . For instance,  $\lfloor 6.2 \rfloor = 6$  and  $\lfloor -1.66 \rfloor = -2$ .

Similarly, there exists a *ceiling functor*  $\text{Ceil} : \mathbb{R} \rightarrow \mathbb{Z}$  that maps each real number  $r$  to the least integer greater than or equal to  $r$ , denoted  $\lceil r \rceil$ .

**Definition 2.3.35.** Given categories  $C$ ,  $D$ , and  $E$ , a *bifunctor*  $F : C \times D \rightarrow E$  is simply a functor from the product category  $C \times D$  to  $E$ . In particular,  $F$  is a rule that assigns:

- to every object  $A \in C$  and  $B \in D$ , an object  $F(A, B) \in E$ ;
- to every morphism  $f : A \rightarrow A'$  in  $C$  and  $g : B \rightarrow B'$  in  $D$ , a morphism  $F(f, g) : F(A, B) \rightarrow F(A', B') \in E$ .

These assignments must satisfy the following two requirements:

- **Respect for composition:** For morphisms  $f : A \rightarrow A'$ ,  $f' : A' \rightarrow A''$  in  $C$  and  $g : B \rightarrow B'$ ,  $g' : B' \rightarrow B''$  in  $D$ , it should hold that

$$F(f' \circ f, g' \circ g) = F(f', g') \circ F(f, g),$$

where the  $\circ$  on the right-hand side is composition in  $E$ .

- **Respect for identities:** For all objects  $A \in C$  and  $B \in D$ , it should hold that

$$F(\text{id}_A, \text{id}_B) = \text{id}_{F(A, B)},$$

where  $\text{id}_A$  and  $\text{id}_B$  are the identity morphisms in  $C$  and  $D$ , respectively, and  $\text{id}_{F(A, B)}$  is the identity morphism in  $E$ .

Many times, rather than writing the name of the bifunctor before the input, like  $F(A, B)$ , we write the bifunctor as an operation between the inputs, for example,  $a \square b$ . If we use this notation, the condition

$$F(f' \circ f, g' \circ g) = F(f', g') \circ F(f, g)$$

becomes

$$(f' \circ f) \square (g' \circ g) = (f' \square g') \circ (f \square g).$$

### 2.3.4 Natural Transformations

A *natural transformation* is a morphism between functors. It provides a way of relating two functors that have the same domain and codomain. Intuitively, if we consider two functors  $F, G : C \rightarrow D$  as different ways of assigning images of the category  $C$  into the category  $D$ , then a natural transformation  $\eta : F \Rightarrow G$  is a coherent way of transforming the image of  $F$  into the image of  $G$ .

**Definition 2.3.36.** Let  $C$  and  $D$  be categories, and let  $F, G : C \rightarrow D$  be functors. A *natural transformation*  $\eta : F \Rightarrow G$  is a family of morphisms in  $D$ ,

$$(\eta_A : FA \rightarrow GA)_{A \in \text{Ob}(C)},$$

indexed by the objects of  $C$ , such that for every morphism  $f : A \rightarrow A'$  in  $C$ , the following diagram commutes.

$$\begin{array}{ccc} FA & \xrightarrow{\eta_A} & GA \\ Ff \downarrow & & \downarrow Gf \\ FA' & \xrightarrow{\eta_{A'}} & GA' \times B' \end{array}$$

Given a natural transformation  $\eta : F \Rightarrow G$ , the morphism  $\eta_A : F(A) \rightarrow G(A)$  in  $D$  is called the *component* of  $\eta$  at  $A$ . A natural transformation  $\eta : F \Rightarrow G$  is represented diagrammatically as

$$\begin{array}{ccc} & F & \\ C & \xrightarrow{\quad} & D \\ & G & \end{array} \quad \Downarrow \eta$$

**Example 2.3.37.** For every functor  $F : C \rightarrow D$ , there exists a natural transformation

$$\iota_F : F \Rightarrow F$$

known as *identity natural transformation*, such that for each object  $A \in \mathbf{C}$ , each component of  $\iota_F$  is the identity morphism:

$$(\iota_F)_A = \text{id}_{F(A)} : F(A) \rightarrow F(A).$$

**Example 2.3.38.** The *list functor*

$$\text{List} : \mathbf{Set} \rightarrow \mathbf{Set}$$

assigns to each set  $S$  the set of all finite sequences (or lists) of its elements.

For instance, if  $S = \{a, b, c\}$ , then

$$\text{List}(S) = \{\varepsilon, a, b, c, aa, ab, ac, ba, \dots, abc, cba, \dots\},$$

where  $\varepsilon$  denotes the empty list.

Given a function  $f : S \rightarrow T$ , where  $T = \{1, 2\}$ , the functor maps it to  $\text{List}(f) : \text{List}(S) \rightarrow \text{List}(T)$ , which applies  $f$  to each element of a list. For example, if

$$f(a) = 2, \quad f(b) = 1, \quad f(c) = 2,$$

then  $\text{List}(f)(abccba) = 2212212$ .

There exists a natural transformation

$$\text{Reverse} : \text{List} \Rightarrow \text{List},$$

whose component at a set  $S$ ,  $\text{Reverse}_S$ , maps each list to its reversal. For example:

$$\text{Reverse}_S(accbab) = babcca.$$

**Definition 2.3.39.** A natural transformation  $\eta : F \Rightarrow G$  between functors  $F, G : \mathbf{C} \rightarrow \mathbf{D}$  is called a *natural isomorphism* if, for every object  $A \in \mathbf{C}$ ,  $\eta_A : F(A) \rightarrow G(A)$  is an isomorphism in  $\mathbf{D}$ .

## 2.3.5 Equivalence of Categories

In category theory, the concept of isomorphism between categories can be quite strict - is not often is arises in a non-trivial manner. So, we weaken the notion of isomorphism and come to the concept of an *equivalence of categories*.

If  $F : C \rightarrow D$  is an isomorphism of categories, then for every object  $B \in D$ , there exists a *unique* object  $A \in C$  such that  $F(A) = B$ . This expresses the idea that  $C$  and  $D$  are structurally identical.

An *equivalence of categories* relaxes this requirement. For every object  $B \in D$ , there exists an object  $A \in C$  such that  $F(A)$  is not necessarily equal to  $B$ , but is *isomorphic* to  $B$ .

**Definition 2.3.40.** Categories  $C$  and  $D$  are said to be *equivalent* if there exist functors  $F : C \rightarrow D$  and  $G : D \rightarrow C$  such that  $G \circ F \cong \text{id}_C$  and  $F \circ G \cong \text{id}_D$ . The functors  $F$  and  $G$  are called *quasi-inverses*, and we write  $C \simeq D$ . This means that for every  $A \in C$ , there is a  $B \in D$  with  $G(B) \cong A$ , and for every  $B \in D$ , there is an  $A \in C$  with  $F(A) \cong B$ .

**Example 2.3.41.** One of the simplest examples of an equivalence of categories is the relationship between the one-object category  $\mathbf{1}$  and the category  $\mathbf{2}_I$ , which has two objects and a single isomorphism between them. We can visualize this as:

$$* \quad \simeq \quad a \xrightarrow{\cong} b$$

More precisely, there is a unique functor  $! : \mathbf{2}_I \rightarrow \mathbf{1}$ , and a functor  $L : \mathbf{1} \rightarrow \mathbf{2}_I$  defined by  $L(*) = a$ . Clearly, the composition  $! \circ L$  is equal to  $\text{id}_{\mathbf{1}}$ , and  $L \circ ! \cong \text{id}_{\mathbf{2}_I}$ , since both objects  $a$  and  $b$  in  $\mathbf{2}_I$  are isomorphic. Thus,  $\mathbf{1} \simeq \mathbf{2}_I$ .

### 2.3.6 Adjoints

If we further weaken the notion of an equivalence of categories, we arrive at the concept of an *adjunction*. A central idea in category theory is that *the weaker the assumptions we impose, the more mathematical phenomena we can capture*. This principle explains why adjunctions are among category theory's most important structures.

**Definition 2.3.42.** An adjunction between categories  $C$  and  $D$  is given by a quadruple  $(L, R, \eta, \epsilon)$ , where  $L : C \rightarrow D$  and  $R : D \rightarrow C$  are functors, and  $\eta : \text{id}_C \Rightarrow RL$  and  $\epsilon : LR \Rightarrow \text{id}_D$  are natural transformations such that

$$(\eta R) \circ (R \epsilon) = \text{id}_R \quad \text{and} \quad (L \eta) \circ (\epsilon L) = \text{id}_L.$$

One says that  $R$  is right adjoint to  $L$ , or that  $L$  is left adjoint to  $R$ , and one calls  $\eta$  the *unit* and  $\epsilon$  the *counit* of the adjunction. Such an adjunction is denoted by  $L \dashv R$ , where the turn of the symbol  $\dashv$  always points to the left adjoint.

An adjunction between categories can be characterized in multiple equivalent ways. It follows one of the alternative formulations.

Given categories  $\mathcal{C}$  and  $\mathcal{D}$ , a pair of functors  $L : \mathcal{C} \rightarrow \mathcal{D}$  and  $R : \mathcal{D} \rightarrow \mathcal{C}$  form an *adjunction*  $L \dashv R$  if there exists a natural isomorphism:

$$\text{Hom}_{\mathcal{D}}(L(A), B) \xrightarrow{\Phi_{A,B}} \text{Hom}_{\mathcal{C}}(A, R(B)).$$

**Example 2.3.43.** Consider the set of real numbers  $\mathbb{R}$  and the set of integers  $\mathbb{Z}$ , each viewed as partial order categories. There is an inclusion functor  $\text{inc} : \mathbb{Z} \hookrightarrow \mathbb{R}$  which simply maps each integer to itself. This inclusion has a left adjoint  $L : \mathbb{R} \rightarrow \mathbb{Z}$ .

To determine this left adjoint  $L$ , we use the definition of an adjunction: for all  $N \in \mathbb{Z}$  and  $R \in \mathbb{R}$ , we have a natural isomorphism:

$$\text{Hom}_{\mathbb{Z}}(L(R), N) \cong \text{Hom}_{\mathbb{R}}(R, \text{inc}(N)).$$

Since both  $\mathbb{Z}$  and  $\mathbb{R}$  are partial orders, the hom-sets contain at most one morphism. Hence, this isomorphism reduces to the logical equivalence:

$$L(R) \leq N \text{ if and only if } R \leq \text{inc}(N) = N.$$

Take  $R = 7.27$  as an example. Then the inequality  $R \leq N$  holds precisely when  $N$  is an integer greater than or equal to 7.27. That is:

$$7.27 \not\leq 5, \quad 7.27 \not\leq 6, \quad 7.27 \not\leq 7, \quad 7.27 \leq 8, \quad 7.27 \leq 9, \quad \dots$$

By the condition above, we must then have:

$$L(7.27) \not\leq 5, \quad L(7.27) \not\leq 6, \quad L(7.27) \not\leq 7, \quad L(7.27) \leq 8, \quad L(7.27) \leq 9, \quad \dots$$

From this, we conclude that  $L(7.27) = 8$ . In general,  $L(R)$  is the least integer greater than or equal to  $R$ , i.e., the ceiling function:

$$L(r) = \lceil r \rceil.$$

Thus, the inclusion functor  $\text{inc}$  has  $\lceil \cdot \rceil$  as a left adjoint, i.e.,  $\lceil \cdot \rceil \dashv \text{inc}$ . The unit of this adjunction is the natural transformation  $\eta : \text{id}_{\mathbb{R}} \Rightarrow \text{inc} \circ \lceil \cdot \rceil$ , which expresses the inequality  $r \leq \lceil r \rceil$  for all  $r \in \mathbb{R}$ . The counit of the adjunction is the identity, since for any integer  $n$ , it holds that  $\lceil n \rceil = n$ .

**Definition 2.3.44.** Let  $F : \mathcal{C} \rightarrow \mathcal{D}$  and  $G : \mathcal{D} \rightarrow \mathcal{E}$  be functors. It is said that  $G$  *preserves coproducts* if whenever  $L$  is a coproduct of  $F$ , then  $G(L)$  is a coproduct of  $G \circ F$ .

**Theorem 2.3.45.** *Left adjoints preserve coproducts.*

### 2.3.7 Monoidal categories

**Definition 2.3.46.** A **monoid** is a triple  $(M, \cdot, u)$ , where  $M$  is a set equipped with a binary operation  $\cdot : M \times M \rightarrow M$  and a distinguished element  $u \in M$  called the *unit*, satisfying the following axioms for all  $x, y, z \in M$ :

$$\text{(Associativity)} \quad x \cdot (y \cdot z) = (x \cdot y) \cdot z,$$

$$\text{(Unit laws)} \quad u \cdot x = x = x \cdot u.$$

Monoidal categories are named so because they are categories equipped with an additional structure that resembles the structure of monoids.

**Definition 2.3.47.** A symmetric monoidal category consists of a category  $\mathcal{C}$  equipped with a bifunctor  $\otimes : \mathcal{C} \times \mathcal{C} \rightarrow \mathcal{C}$  called *tensor product* and a distinguished object  $I \in \mathcal{C}$ , called *unit* together with natural isomorphisms

$$\alpha_{A,B,C} : A \otimes (B \otimes C) \rightarrow (A \otimes B) \otimes C,$$

$$\lambda_A : I \otimes A \rightarrow A, \quad \rho_A : A \otimes I \rightarrow A,$$

known as *reassociator*, *left unitor*, and *right unitor*, respectively.

Moreover, these natural isomorphisms are required to make the following coherence diagrams commute.

$$\begin{array}{ccc} & (A \otimes B) \otimes (C \otimes D) & \\ \alpha \nearrow & & \searrow \alpha \\ A \otimes (B \otimes (C \otimes D)) & & ((A \otimes B) \otimes C) \otimes D \\ \text{id} \otimes \alpha \downarrow & & \uparrow \alpha \otimes \text{id} \\ A \otimes ((B \otimes C) \otimes D) & \xrightarrow{\alpha} & (A \otimes (B \otimes C)) \otimes D \end{array}$$
  

$$\begin{array}{ccc} A \otimes (I \otimes A) & \xrightarrow{\alpha} & (A \otimes I) \otimes A \\ \text{id} \otimes \lambda_A \searrow & & \nearrow \rho_A \otimes \text{id} \\ & A \otimes A & \end{array} \quad \begin{array}{ccc} I \otimes I & \xlongequal{\quad} & I \otimes I \\ \lambda_I \searrow & & \nearrow \rho_I \\ & I & \end{array}$$

**Definition 2.3.48.** A monoidal category is said to be *symmetric* when it is equipped with a natural isomorphism  $\text{sw} : A \otimes B \rightarrow B \otimes A$  known as *braiding* such that the following diagrams commute.

$$\begin{array}{ccc}
 A \otimes B & \xrightarrow{\text{sw}_{A,B}} & B \otimes A \\
 & \searrow & \swarrow \\
 & A \otimes B &
 \end{array}
 \qquad
 \begin{array}{ccc}
 A \otimes \mathbb{I} & \xrightarrow{\text{sw}_{A,\mathbb{I}}} & \mathbb{I} \otimes A \\
 & \searrow \rho_A & \swarrow \lambda_A \\
 & A &
 \end{array}$$
  

$$\begin{array}{ccccc}
 A \otimes (B \otimes C) & \xrightarrow{\alpha} & (A \otimes B) \otimes C & \xrightarrow{\text{sw}} & C \otimes (A \otimes B) \\
 \downarrow \text{id} \otimes \text{sw} & & & & \downarrow \alpha \\
 A \otimes (C \otimes B) & \xrightarrow{\alpha} & (A \otimes C) \otimes B & \xrightarrow{\text{sw} \otimes \text{id}} & (C \otimes A) \otimes B
 \end{array}$$

**Definition 2.3.49.** A monoidal category  $\mathcal{C}$  is said to be *closed* if for each object  $A$  in  $\mathcal{C}$  the functor  $- \otimes A$  has a right adjoint, denoted by  $A \multimap -$ .

**Definition 2.3.50.** A monoidal category  $\mathcal{C}$  with coproducts is called *distributive* if for every object  $A$  in  $\mathcal{C}$  the functor  $- \otimes A$  preserves coproducts. Explicitly this means that the morphism,

$$[\text{inl} \otimes \text{id}, \text{inr} \otimes \text{id}] : B \otimes A \oplus C \otimes A \rightarrow (B \oplus C) \otimes A$$

is actually an isomorphism. We will denote the respective inverse by  $\text{dist}$ . Note that if  $\mathcal{C}$  is monoidal closed then it is automatically distributive as left adjoints preserve all colimits.

**Definition 2.3.51.** The *terminal map to the unit object* (of a monoidal category) is the unique mapping from each object to the monoidal unit and is represented by  $!$ .

**Example 2.3.52.** Examples of monoidal closed categories with coproducts include  $\text{Set}$  and  $\text{CVect}$ . In  $\text{Set}$ , the tensor product is the cartesian product, the monoidal unit is the singleton set, the coproduct is the disjoint union, and the internal hom consists of all functions between sets. For  $\text{CVect}$ , the tensor product is the standard tensor product of complex vector spaces, the field of complex number  $\mathbb{C}$ , the coproduct is the direct sum, and the internal hom is the space of complex linear maps.



**Theorem 2.3.53** (*Coherence Theorem for Symmetric Monoidal Categories*). Any diagram in a symmetric monoidal category constructed only from associators  $\alpha$ , unitors  $\lambda, \rho$ , the symmetry  $sw$ , and inverses and their composition and tensor product necessarily commutes.

## 2.4 Interpretation

Sets, sub, exchange, soundness and completeness, conds, diagrama a explicar

Up to this point, we have discussed  $\lambda$ -calculus in abstract terms: we explored which programs can be written, but we have not yet assigned them any meaning. This process—assigning meaning to syntactic expressions—is known as the *interpretation* or *semantics* of the language. In fact, the word “semantics” comes from the Greek word for “meaning”.

There are different kinds of semantics, in particular, *denotational semantics* interprets terms as mathematical objects. This is done by defining a function that maps syntactic entities (such as types and terms) to semantic entities (such as sets and elements). This mapping is called the *interpretation function*, typically denoted by  $\llbracket - \rrbracket$ . Thus, given a term  $v$ , we write  $\llbracket v \rrbracket$  to denote its meaning under a specific interpretation.

Naturally, this raises important questions: what guarantees do we have that the interpretation of terms respects the classical equations of the calculus? This leads us to the notions of *soundness* and *completeness*.

With respect to a given class of interpretations:

- *Soundness* is the property

$$M = N \Rightarrow \llbracket M \rrbracket = \llbracket N \rrbracket \quad \text{for all interpretations in the class.}$$

That is, if two terms are provably equal, then they are interpreted as equal.

- *Completeness* is the property

$$\llbracket M \rrbracket = \llbracket N \rrbracket \Rightarrow M = N \quad \text{for all interpretations in the class.}$$

That is, if two terms are interpreted as equal, then they are provably equal.

Soundness ensures that our equations are *correct*—all derivable equations are semantically valid. Completeness ensures that our equations are *sufficient*—we can derive all semantically valid equations.

We note that, in the case of the metric equations, the underlying idea is similar, although soundness and completeness are defined differently.

In order to define the interpretation of judgments  $\Gamma \triangleright v : \mathbb{A}$ , it is necessary to establish some notation first. Let  $\mathbf{C}$  be a symmetric monoidal category and  $A, B$  and  $C$  be objects of this category. For all morphisms  $f : A \otimes B \rightarrow C$ , the morphism  $\bar{f} : A \rightarrow (B \multimap C)$  denotes the corresponding curried version. Moreover, there is the application morphism,  $\text{app} : (A \multimap B) \otimes A \rightarrow B$ .

For all ground types  $X \in G$  the interpretation of  $\llbracket X \rrbracket$  is postulated to be an object of  $\mathbf{C}$ . Types are interpreted inductively using the unit  $\mathbb{I}$ , the tensor  $\otimes$ , the coproduct  $\oplus$ , and the linear map  $\multimap$ . Given a non-empty context  $\Gamma = \Gamma', x : \mathbb{A}$ , its interpretation is defined by  $\llbracket \Gamma', x : \mathbb{A} \rrbracket = \llbracket \Gamma' \rrbracket \otimes \llbracket \mathbb{A} \rrbracket$  if  $\Gamma'$  is non-empty and  $\llbracket \Gamma', x : \mathbb{A} \rrbracket = \llbracket \mathbb{A} \rrbracket$  otherwise. The empty context  $-$  is interpreted as  $\llbracket - \rrbracket = \mathbb{I}$ . Given  $A_1, \dots, A_n \in \mathbf{C}$ , the  $n$ -tensor  $(\dots (A_1 \otimes A_2) \otimes \dots) \otimes A_n$  is denoted as  $X_1 \otimes \dots \otimes A_n$ , and similarly for morphisms.

### 2.4.1 Classical Semantics

“Housekeeping” morphisms are employed to handle interactions between context interpretation and the vectorial model. Given  $\Gamma_1, \dots, \Gamma_n$ , the morphism that splits  $\llbracket \Gamma_1, \dots, \Gamma_n \rrbracket$  into  $\llbracket \Gamma_1 \rrbracket \otimes \dots \otimes \llbracket \Gamma_n \rrbracket$  is denoted by  $\text{sp}_{\Gamma_1, \dots, \Gamma_n} : \llbracket \Gamma_1, \dots, \Gamma_n \rrbracket \rightarrow \llbracket \Gamma_1 \rrbracket \otimes \dots \otimes \llbracket \Gamma_n \rrbracket$ . For  $n = 1$ ,  $\text{sp}_{\Gamma_1} = \text{id}$ . Let  $\Gamma_1$  and  $\Gamma_2$  be two contexts,  $\text{sp}_{\Gamma_1, \Gamma_2} : \Gamma_1 \otimes \Gamma_2$  is defined as:

$$\text{sp}_{-; \Gamma_2} = \lambda^{-1} \quad \text{sp}_{\Gamma_1; -} = \rho^{-1} \quad \text{sp}_{\Gamma_1; x : \mathbb{A}} = \text{id} \quad \text{sp}_{\Gamma_1; \Delta, x : \mathbb{A}} = \alpha \cdot (\text{sp}_{\Gamma_1; \Delta} \otimes \text{id})$$

For  $n > 2$ ,  $\text{sp}_{\Gamma_1, \dots, \Gamma_n}$  is defined recursively based on the previous definition, using induction on  $n$ :

$$\text{sp}_{\Gamma_1, \dots, \Gamma_n} = (\text{sp}_{\Gamma_1, \dots, \Gamma_{n-1}} \otimes \text{id}) \cdot \text{sp}_{\Gamma_1, \dots, \Gamma_{n-1}; \Gamma_n}$$

On the other hand,  $\text{jn}_{\Gamma_1, \dots, \Gamma_n}$  denotes the inverse of  $\text{sp}_{\Gamma_1, \dots, \Gamma_n}$ . Next, given  $\Gamma, x : \mathbb{A}, y : \mathbb{B}, \Delta$ , the morphism permuting  $x$  and  $y$  is denoted by  $\text{exch}_{\Gamma, x : \mathbb{A}, y : \mathbb{B}, \Delta} : \llbracket \Gamma, x : \mathbb{A}, y : \mathbb{B}, \Delta \rrbracket \rightarrow \llbracket \Gamma, y : \mathbb{B}, x : \mathbb{A}, \Delta \rrbracket$  and defined as:

$$\text{exch}_{\Gamma, x : \mathbb{A}, y : \mathbb{B}, \Delta} = \text{jn}_{\Gamma; y : \mathbb{B}, x : \mathbb{A}; \Delta} \cdot (\text{id} \otimes \text{sw} \otimes \text{id}) \cdot \text{sp}_{\Gamma; x : \mathbb{A}, y : \mathbb{B}; \Delta}$$

The shuffling morphism  $\text{sh}_E : \llbracket E \rrbracket \rightarrow \llbracket \Gamma_1, \dots, \Gamma_n \rrbracket$  is defined as a suitable composition of exchange morphisms.

For every operation symbol  $f : \mathbb{A}_1, \dots, \mathbb{A}_n \rightarrow \mathbb{A}$  it is assumed the existence of an morphism  $\llbracket f \rrbracket : \llbracket \mathbb{A}_1 \rrbracket \otimes \dots \otimes \llbracket \mathbb{A}_n \rrbracket \rightarrow \llbracket \mathbb{A} \rrbracket$ . The interpretation of judgments is defined by induction over derivations according to the rules in [Figure 4](#).

$$\begin{array}{c}
\frac{\llbracket \Gamma_i \triangleright v_i : \mathbb{A}_i \rrbracket = m_i \quad f : \mathbb{A}_1, \dots, \mathbb{A}_n \rightarrow \mathbb{A} \in \Sigma \quad E \in \mathbf{Sf}(\Gamma_1; \dots; \Gamma_n)}{\llbracket E \triangleright f(v_1, \dots, v_n) : \mathbb{A} \rrbracket = \llbracket f \rrbracket \cdot (m_1 \otimes \dots \otimes m_n) \cdot \mathbf{sp}_{\Gamma_1; \dots; \Gamma_n} \cdot \mathbf{sh}_E} \quad \llbracket x : \mathbb{A} \triangleright x : \mathbb{A} \rrbracket = \text{id}_{\llbracket \mathbb{A} \rrbracket} \\
\\
\frac{\llbracket \Gamma \triangleright v : \mathbb{A} \otimes \mathbb{B} \rrbracket = m \quad \llbracket \Delta, x : \mathbb{A}, y : \mathbb{B} \triangleright w : \mathbb{D} \rrbracket = n \quad E \in \mathbf{Sf}(\Gamma; \Delta)}{\llbracket - \triangleright * : \mathbb{I} \rrbracket = \text{id}_{\llbracket \mathbb{I} \rrbracket} \quad \llbracket E \triangleright \mathbf{pm} \, v \, \mathbf{to} \, x \otimes y.w : \mathbb{D} \rrbracket = n \cdot \mathbf{jn}_{\Delta; \mathbb{A}; \mathbb{B}} \cdot \alpha \cdot \mathbf{sw} \cdot (m \otimes \text{id}) \cdot \mathbf{sp}_{\Gamma; \Delta} \cdot \mathbf{sh}_E} \\
\\
\frac{\llbracket \Gamma \triangleright v : \mathbb{A} \rrbracket = m \quad \llbracket \Delta \triangleright w : \mathbb{B} \rrbracket = n \quad E \in \mathbf{Sf}(\Gamma; \Delta)}{\llbracket E \triangleright v \otimes w : \mathbb{A} \otimes \mathbb{B} \rrbracket = (m \otimes n) \cdot \mathbf{sp}_{\Gamma; \Delta} \cdot \mathbf{sh}_E} \\
\\
\frac{\llbracket \Gamma \triangleright v : \mathbb{I} \rrbracket = m \quad \llbracket \Delta \triangleright w : \mathbb{A} \rrbracket = n \quad E \in \mathbf{Sf}(\Gamma; \Delta)}{\llbracket E \triangleright v \, \mathbf{to} \, * . w : \mathbb{A} \rrbracket = n \cdot \lambda \cdot (m \otimes \text{id}) \cdot \mathbf{sp}_{\Gamma; \Delta} \cdot \mathbf{sh}_E} \quad \frac{\llbracket \Gamma, x : \mathbb{A} \triangleright v : \mathbb{B} \rrbracket = m}{\llbracket \Gamma \triangleright \lambda x : \mathbb{A}. v : \mathbb{A} \multimap \mathbb{B} \rrbracket = \overline{m \cdot \mathbf{jn}_{\Gamma; \mathbb{A}}}} \\
\\
\frac{\llbracket \Gamma \triangleright v : \mathbb{A} \multimap \mathbb{B} \rrbracket = m \quad \llbracket \Delta \triangleright w : \mathbb{A} \rrbracket = n \quad E \in \mathbf{Sf}(\Gamma; \Delta)}{\llbracket E \triangleright vw : \mathbb{B} \rrbracket = \mathbf{app} \cdot (m \otimes n) \cdot \mathbf{sp}_{\Gamma; \Delta} \cdot \mathbf{sh}_E} \quad \frac{\llbracket \Gamma \triangleright v : \mathbb{A} \rrbracket = f}{\llbracket \Gamma \triangleright \mathbf{dis}(v) : \mathbb{I} \rrbracket = \mathbf{!} \cdot f} \\
\\
\frac{\llbracket \Gamma \triangleright v : \mathbb{A} \rrbracket = m}{\llbracket \Gamma \triangleright \mathbf{inl}_{\mathbb{B}}(v) : \mathbb{A} \oplus \mathbb{B} \rrbracket = \mathbf{inl} \cdot m} \quad \frac{\llbracket \Gamma \triangleright v : \mathbb{B} \rrbracket = m}{\llbracket \Gamma \triangleright \mathbf{inr}_{\mathbb{A}}(v) : \mathbb{A} \oplus \mathbb{B} \rrbracket = \mathbf{inr} \cdot m} \\
\\
\frac{\llbracket \Gamma \triangleright v : \mathbb{A} \oplus \mathbb{B} \rrbracket = b \quad \llbracket \Delta, x : \mathbb{A} \triangleright w : \mathbb{D} \rrbracket = p \quad \llbracket \Delta, y : \mathbb{B} \triangleright u : \mathbb{D} \rrbracket = q \quad E \in \mathbf{Sf}(\Gamma; \Delta)}{\llbracket E \triangleright \mathbf{case} \, v \, \{ \mathbf{inl}_{\mathbb{B}}(x) \Rightarrow w; \mathbf{inr}_{\mathbb{A}}(y) \Rightarrow u \} : \mathbb{D} \rrbracket = [p, q] \cdot (\mathbf{jn}_{\Delta; \mathbb{A}} \cdot \mathbf{sw} \oplus \mathbf{jn}_{\Delta; \mathbb{B}} \cdot \mathbf{sw}) \cdot \mathbf{dist} \cdot (b \otimes \text{id}) \cdot \mathbf{sp}_{\Gamma; \Delta} \cdot \mathbf{sh}_E}
\end{array}$$

Figure 4: Judgment interpretation

The following diagrams are useful for a clearer understanding of the interpretation of judgments given in [Figure 4](#).

$$\begin{array}{ll}
\llbracket \mathbf{ax} \rrbracket : & \llbracket E \rrbracket \xrightarrow{\mathbf{sh}_E} \llbracket \Gamma_1, \dots, \Gamma_n \rrbracket \xrightarrow{\mathbf{sp}_{\Gamma; \Delta}} \llbracket \Gamma_1 \rrbracket \otimes \dots \otimes \llbracket \Gamma_n \rrbracket \\
& \xrightarrow{m_1 \otimes \dots \otimes m_n} \llbracket \mathbb{A}_1 \rrbracket \otimes \dots \otimes \llbracket \mathbb{A}_n \rrbracket \xrightarrow{\llbracket f \rrbracket} \llbracket \mathbb{A} \rrbracket \\
\llbracket \mathbf{hyp} \rrbracket : & \llbracket \mathbb{A} \rrbracket \xrightarrow{\text{id}_{\llbracket \mathbb{A} \rrbracket}} \llbracket \mathbb{A} \rrbracket \\
\llbracket \mathbb{I}_i \rrbracket : & \llbracket \mathbb{I} \rrbracket \xrightarrow{\text{id}_{\llbracket \mathbb{I} \rrbracket}} \llbracket \mathbb{I} \rrbracket \\
\llbracket \otimes_e \rrbracket : & \llbracket E \rrbracket \xrightarrow{\mathbf{sh}_E} \llbracket \Gamma, \Delta \rrbracket \xrightarrow{\mathbf{sp}_{\Gamma; \Delta}} \llbracket \Gamma \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{m \otimes \text{id}} (\llbracket \mathbb{A} \rrbracket \otimes \llbracket \mathbb{B} \rrbracket) \otimes \llbracket \Delta \rrbracket \\
& \xrightarrow{\mathbf{sw}} \llbracket \Delta \rrbracket \otimes (\llbracket \mathbb{A} \rrbracket \otimes \llbracket \mathbb{B} \rrbracket) \xrightarrow{\alpha} (\llbracket \Delta \rrbracket \otimes \llbracket \mathbb{A} \rrbracket) \otimes \llbracket \mathbb{B} \rrbracket \xrightarrow{\mathbf{jn}_{\Delta; \mathbb{A}; \mathbb{B}}} \llbracket \Delta, \mathbb{A}, \mathbb{B} \rrbracket \\
& \xrightarrow{n} \llbracket \mathbb{D} \rrbracket
\end{array}$$

$$\begin{aligned}
\llbracket \otimes_i \rrbracket : \quad & \llbracket E \rrbracket \xrightarrow{sh_E} \llbracket \Gamma, \Delta \rrbracket \xrightarrow{sp_{\Gamma, \Delta}} \llbracket \Gamma \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{m \otimes n} \llbracket A \rrbracket \otimes \llbracket B \rrbracket \\
\llbracket \mathbb{I}_e \rrbracket : \quad & \llbracket E \rrbracket \xrightarrow{sh_E} \llbracket \Gamma, \Delta \rrbracket \xrightarrow{sp_{\Gamma, \Delta}} \llbracket \Gamma \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{m \otimes id} \llbracket \mathbb{I} \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{\lambda} \llbracket \Delta \rrbracket \xrightarrow{n} \llbracket A \rrbracket \\
\llbracket -\circ_i \rrbracket : \quad & \llbracket \Gamma \rrbracket \otimes \llbracket A \rrbracket \xrightarrow{\overline{m \cdot jn_{\Gamma, A}}} \llbracket A \rrbracket \multimap \llbracket B \rrbracket \quad (\llbracket \Gamma \rrbracket \otimes \llbracket A \rrbracket \xrightarrow{jn_{\Gamma, A}} \llbracket \Gamma, A \rrbracket \xrightarrow{m} \llbracket B \rrbracket) \\
\llbracket -\circ_e \rrbracket : \quad & \llbracket E \rrbracket \xrightarrow{sh_E} \llbracket \Gamma, \Delta \rrbracket \xrightarrow{sp_{\Gamma, \Delta}} \llbracket \Gamma \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{m \otimes n} \llbracket A \rrbracket \multimap \llbracket B \rrbracket \otimes \llbracket A \rrbracket \xrightarrow{app} \llbracket B \rrbracket \\
\llbracket dis \rrbracket : \quad & \llbracket \Gamma \rrbracket \xrightarrow{f} \llbracket A \rrbracket \xrightarrow{l} \llbracket \mathbb{I} \rrbracket \\
\llbracket inl \rrbracket : \quad & \llbracket \Gamma \rrbracket \xrightarrow{m} \llbracket A \rrbracket \xrightarrow{inl} \llbracket A \oplus B \rrbracket \\
\llbracket inr \rrbracket : \quad & \llbracket \Gamma \rrbracket \xrightarrow{m} \llbracket B \rrbracket \xrightarrow{inr} \llbracket A \oplus B \rrbracket \\
\llbracket case \rrbracket : \quad & \llbracket E \rrbracket \xrightarrow{sh_E} \llbracket \Gamma, \Delta \rrbracket \xrightarrow{sp_{\Gamma, \Delta}} \llbracket \Gamma \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{b \otimes id} (\llbracket A \rrbracket \oplus \llbracket B \rrbracket) \otimes \llbracket \Delta \rrbracket \\
& \xrightarrow{dist} (\llbracket A \rrbracket \otimes \llbracket \Delta \rrbracket) \oplus (\llbracket B \rrbracket \otimes \llbracket \Delta \rrbracket) \\
& \xrightarrow{jn_{\Delta, A} \cdot sw \oplus jn_{\Delta, B} \cdot sw} (\llbracket \Delta, A \rrbracket) \oplus (\llbracket \Delta, B \rrbracket) \xrightarrow{[p, q]} \llbracket D \rrbracket
\end{aligned}$$

Regarding the interpretation of the exchange and substitution properties, we have the following lemma.

**Lemma 2.4.1.** *For any judgements  $\Gamma, x : A, y : B, \Delta \triangleright v : D, \Gamma, x : A \triangleright v : B$ , and  $\Delta \triangleright w : A$ , the following holds:*

$$\begin{aligned}
\llbracket \Gamma, x : A, y : B, \Delta \triangleright v : D \rrbracket &= \llbracket \Gamma, y : B, x : A, \Delta \triangleright v : D \rrbracket \cdot \text{exch}_{\Gamma, x:A, y:B, \Delta} \\
\llbracket \Gamma, \Delta \triangleright v[w/x] : B \rrbracket &= \llbracket \Gamma, x : A \triangleright v : B \rrbracket \cdot jn_{\Gamma, A} \cdot (\text{id} \otimes \llbracket \Delta \triangleright w : A \rrbracket) \cdot sp_{\Gamma, \Delta}
\end{aligned}$$

*Proof.* This lemma is proved in [24, Lemma 2.2] for the lambda calculus without conditionals, so we only need to address the conditional cases.

Passar para aqui as provas qd elas estiverem vistas

□

**Theorem 2.4.2.** *The equations presented in Figure 2 are sound with respect to judgement interpretation. More specifically, if  $\Gamma \triangleright v = w : A$  is one of the equations in Figure 2, then  $\llbracket \Gamma \triangleright v : A \rrbracket = \llbracket \Gamma \triangleright w : A \rrbracket$ .*

*Proof.* Given the theorem is already proven in [24, Theorem 2.3] for the lambda calculus without conditionals, it suffices to consider the cases involving conditionals.

Passar para aqui as provas qd elas estiverem vistas

□

**Definition 2.4.3** (Models of linear  $\lambda$ -theories). Consider a linear  $\lambda$ -theory  $((G, \Sigma), Ax)$  and a symmetric monoidal closed category with coproducts  $\mathbf{C}$ . Suppose that for each  $X \in G$ , we have an interpretation  $\llbracket X \rrbracket$ , which is an object of  $\mathbf{C}$ , and analogously for the operation symbols in  $\Sigma$ . This interpretation structure is a *model* of the theory if all axioms in  $Ax$  are satisfied by the interpretation.

**Theorem 2.4.4** (Completeness). Consider a linear  $\lambda$ -theory  $\mathcal{T}$ . Then an equation  $\Gamma \triangleright v = w : \mathbb{A}$  is a theorem of  $\mathcal{T}$  if and only if it is satisfied by all models of the theory.

*Proof.* This theorem is proved in [24, Lemma 2.6] for the lambda calculus without conditionals, so we only need to address the conditional cases.

Passar para aqui as provas qd elas estiverem vistas

□

**Example 2.4.5.** We now illustrate how the programs presented in Examples 2.2.1, 2.2.2, and 2.2.3, with slight modifications, are interpreted in  $\mathbf{Set}$ . To this effect, we consider a type  $N$  representing set of natural numbers, along with a family of operations  $\{n : \mathbb{I} \rightarrow N \mid n \in \mathbb{N}\}$ , each mapping the monoidal unit to a corresponding natural number  $n$ . In  $\mathbf{Set}$  have  $\llbracket \mathbb{I} \rrbracket = \{*\}$ , and define  $\llbracket N \rrbracket = \mathbb{N}$ , and  $\llbracket n \rrbracket = \{*\} \rightarrow \mathbb{N}, * \mapsto n$ . Consider the following  $\lambda$ -term:

$$x : N \otimes N \triangleright \text{pm } x \text{ to } a \otimes b. b \otimes a [1(*) \otimes 2(*)/x] : N \otimes N$$

Attending to Figure 4 this program is interpreted as follows:

$$\begin{aligned}
& \llbracket \text{pm } x \text{ to } a \otimes b. b \otimes a [1(*) \otimes 2(*)/x] \rrbracket \\
&= \llbracket \text{pm } x \text{ to } a \otimes b. b \otimes a \rrbracket \cdot \text{jn} \cdot (\text{id} \otimes \llbracket 1(*) \otimes 2(*) \rrbracket) \cdot \text{sp} && (\text{Lemma 2.4.1}) \\
&= \llbracket b \otimes a \rrbracket \cdot \text{jn} \cdot \alpha \cdot \text{sw} \cdot (\llbracket x \rrbracket \otimes \text{id}) \cdot \text{sp} \cdot \text{sh} \cdot \lambda \cdot (\text{id} \otimes (\llbracket 1(*) \rrbracket \otimes \llbracket 2(*) \rrbracket) && (\otimes_e, \otimes_i) \\
&\quad \cdot \text{sp} \cdot \text{sh})) \cdot \lambda^{-1} \\
&= \llbracket b \rrbracket \otimes \llbracket a \rrbracket \cdot \text{sp} \cdot \text{sh} \cdot (\lambda \otimes \text{id}) \cdot \alpha \cdot \text{sw} \cdot \rho^{-1} \cdot \lambda \cdot (\text{id} \otimes (1 \cdot \llbracket * \rrbracket) \cdot \text{sp} \cdot \text{sh} \otimes && (\otimes_i, \text{hyp}, \\
&\quad 2 \cdot \llbracket * \rrbracket \cdot \text{sp} \cdot \text{sh} \cdot \lambda^{-1})) \cdot \lambda^{-1} && \text{ax}) \\
&= \text{sw} \cdot \lambda \cdot (\text{id} \otimes (1 \otimes 2 \cdot \lambda^{-1})) \cdot \lambda^{-1} && (\text{hyp}, \mathbb{I}_i, \text{coh}) \\
&= 2 \otimes 1 \cdot \rho^{-1} && (\text{nat}, \text{coh})
\end{aligned}$$

Next, consider the  $\lambda$ -terms below.

$$\mathbf{Dis2nd} \triangleq - \triangleright \lambda x : N \otimes N. \text{pm } x \text{ to } a \otimes b. \text{dis}(b) \text{ to } * . a : N \otimes N \multimap N$$

$$\mathbf{Dis2nd} (1(*) \otimes 2(*))$$

In this case, we will “cheat” slightly by first using the equations in [Figure 2](#) to simplify the term. We note that here we will not explicitly illustrate the interpretation of the case statement, because in  $\mathbf{Set}$ , any term of type  $\mathbb{A} \otimes \mathbb{A}$ , for any type  $\mathbb{A}$ , corresponds to an injection ( $\text{inl}(v)$  or  $\text{inr}(v)$ ). As a result, we can “escape” any case statement by applying the equations  $\beta_{case}^{inl}$  and  $\beta_{case}^{inr}$ .

Applying the  $\beta$  equation we have

$$\begin{aligned} \mathbf{Dis2nd} (1(*) \otimes 2(*)) &= \mathbf{Dis2nd} [1(*) \otimes 2(*)/x] \\ &= \text{pm } 1(*) \otimes 2(*) \text{ to } a \otimes b. \text{dis}(b) \text{ to } *.a \end{aligned}$$

The resulting program is interpreted as follows:

$$\begin{aligned} &\llbracket \text{pm } 1(*) \otimes 2(*) \text{ to } a \otimes b. \text{dis}(b) \text{ to } *.a \rrbracket \\ &= \llbracket \text{dis}(b) \text{ to } *.a \rrbracket \cdot \text{jn} \cdot \alpha \cdot \text{sw} \cdot (\llbracket 1(*) \otimes 2(*) \rrbracket \otimes \text{id}) \cdot \text{sp} \cdot \text{sh} && (\otimes_e) \\ &= \llbracket a \rrbracket \cdot \lambda \cdot (\llbracket \text{dis}(b) \rrbracket \otimes \text{id}) \cdot \text{sp} \cdot \text{sh} \cdot (\lambda \otimes \text{id}) \cdot \alpha \cdot \text{sw} \cdot ((\llbracket 1(*) \rrbracket \otimes \llbracket 2(*) \rrbracket) && (\mathbb{I}_e, \otimes_i) \\ &\quad \cdot \text{sp} \cdot \text{sh}) \otimes \text{id}) \cdot \lambda^{-1} \\ &= \lambda \cdot (! \cdot \llbracket b \rrbracket \otimes \text{id}) \cdot \text{sw} \cdot (\lambda \otimes \text{id}) \cdot \alpha \cdot \text{sw} \cdot ((1 \cdot \llbracket * \rrbracket \cdot \text{sp} \cdot \text{sh} \otimes 2 \cdot \llbracket * \rrbracket \cdot \text{sp} && (\text{hyp}, \text{dis}, \text{ax}) \\ &\quad \cdot \text{sh} \cdot \lambda^{-1}) \otimes \text{id}) \cdot \lambda^{-1} \\ &= \lambda \cdot (! \otimes \text{id}) \cdot \text{sw} \cdot \rho \cdot ((1 \otimes 2 \cdot \lambda^{-1}) \otimes \text{id}) \cdot \lambda^{-1} && (\text{hyp}, \mathbb{I}_i, \text{coh}) \\ &= \lambda \cdot (! \otimes \text{id}) \cdot 2 \otimes 1 \cdot \rho^{-1} && (\text{nat}, \text{coh}) \end{aligned}$$

The diagram below illustrates the resulting interpretation.

$$\{*\} \xrightarrow{\rho^{-1}} \{(*, *)\} \xrightarrow{2 \otimes 1} \{(2, 1)\} \xrightarrow{(! \otimes \text{id})} \{(*, 1)\} \xrightarrow{\lambda} \{1\}$$

## 2.4.2 Semantics of metric equations

**??** We will now turn our attention to the semantics of the metric equations. First, we recall the definition of a metric space.

**Definition 2.4.6.** A *metric space* is a pair  $(X, d)$  where  $X$  is a set and  $d : X \times X \rightarrow [0, \infty)$  is a function known as *distance* satisfying:

1.  $0 \leq d(x, y)$ , with equality if and only if  $x = y$ ,
2.  $d(x, y) = d(y, x)$ ,

3.  $d(x, z) \leq d(x, y) + d(y, z)$  for all  $x, y, z \in X$ .

Here, we equip each hom-set  $C(A, B)$  of a category  $C$  with a metric  $d_{A,B}$ , and impose that both composition and precomposition are non-expansive. That is, for all morphisms  $f, f_1, f_2 \in C(A, B)$  and any  $g, g_1, g_2 \in C(B, C)$ , the following inequalities holds:

$$d_{A,C}(g \circ f_1, g \circ f_2) \leq d_{A,B}(f_1, f_2) \quad d_{A,C}(g_1 \circ f, g_2 \circ f) \leq d_{B,C}(g_1, g_2).$$

Note that, given the triangle inequality, we have:

$$d_{A,C}(g_1 \circ f_1, g_2 \circ f_2) \leq d_{A,C}(g_1 \circ f_1, g_1 \circ f_2) + d_{A,C}(g_1 \circ f_2, g_2 \circ f_2) \leq d_{A,B}(f_1, f_2) + d_{B,C}(g_1, g_2).$$

This is known as *enriching* the category  $C$  over *metric spaces*.

In this context, *soundness* and *completeness* concepts are extended to encompass not only the classical equations but also the metric equations. Note that the  $=$  in classical equations can be written as  $=_0$ . As a result, in this setting we define

- *Soundness* is the property

$$M =_\varepsilon N \Rightarrow d(\llbracket N \rrbracket, \llbracket M \rrbracket) \leq \varepsilon \quad \text{for all interpretations in the class.}$$

That is, if two terms are provably at a maximum distance  $\varepsilon$ , so are their respective interpretations

- *Completeness* is the property

$$d(\llbracket N \rrbracket, \llbracket M \rrbracket) \leq \varepsilon \Rightarrow M =_\varepsilon N \quad \text{for all interpretations in the class.}$$

That is, if  $\varepsilon$  is the maximum distance between the interpretation of two programs, then they are provably at a maximum distance  $\varepsilon$ .

In this context, the notions of *soundness* and *completeness* are extended to encompass not only classical equations but also *metric equations*. Note that classical equations, written as  $M = N$ , can be viewed as a special case of metric equations where the distance is zero, i.e.,  $M =_0 N$ . We define:

- *Soundness* is the property

$$M =_\varepsilon N \Rightarrow d(\llbracket M \rrbracket, \llbracket N \rrbracket) \leq \varepsilon \quad \text{for all interpretations in the class.}$$

That is, if two terms are provably within a distance  $\varepsilon$ , then so are their interpretations.

- *Completeness* is the property

$$d(\llbracket M \rrbracket, \llbracket N \rrbracket) \leq \varepsilon \Rightarrow M =_{\varepsilon} N \quad \text{for all interpretations in the class.}$$

That is, if the interpretations of two terms are within a distance  $\varepsilon$ , then the terms are provably at most a distance  $\varepsilon$  apart.

Esta parte acho que vai ficar um pouco dúbia, melhor confirmar com prof. Renato.

O que fazer com a def de adjunção -> theorem 4.2 artigo

**Definition 2.4.7.** *Met* denotes the category whose objects are metric spaces and whose morphisms are non-expansive maps, *i.e.*, functions that do not increase the distance between points. More precisely, for two metric spaces  $(X, d_X)$  and  $(Y, d_Y)$ , a morphism  $f : (X, d_X) \rightarrow (Y, d_Y)$  is a function  $f : X \rightarrow Y$  such that

$$d_Y(f(x), f(y)) \leq d_X(x, y) \quad \text{for all } x, y \in X.$$

**Definition 2.4.8.** A *Met-category*  $\mathbf{C}$  is a category in which each hom-set  $\mathbf{C}(A, B)$  is equipped with a metric  $d_{A,B}$ , such that for all morphisms  $f_1, f_2 \in \mathbf{C}(A, B)$  and  $g_1, g_2 \in \mathbf{C}(B, C)$ , the following inequality holds:

$$d_{A,C}(g_1 \circ f_1, g_2 \circ f_2) \leq d_{A,B}(f_1, f_2) + d_{B,C}(g_1, g_2).$$

A *symmetric monoidal Met-category*  $\mathbf{C}$  is a category that is both symmetric monoidal and a *Met-category*, such that for all morphisms  $f_1, f_2 \in \mathbf{C}(A, B)$  and  $g_1, g_2 \in \mathbf{C}(C, D)$ , we have:

$$d_{A \otimes C, B \otimes D}(f_1 \otimes g_1, f_2 \otimes g_2) \leq d_{A,B}(f_1, f_2) + d_{C,D}(g_1, g_2).$$

A *symmetric monoidal closed Met-category*  $\mathbf{C}$  is a category that is both symmetric monoidal closed and a symmetric monoidal *Met-category*, such that for all objects  $A, B$ , and  $C$ , there exists a *Met-isomorphism*

$$\mathbf{C}(B \otimes A, C) \cong \mathbf{C}(C, A \multimap B)$$

natural in  $B$  and  $C$ .

**Theorem 2.4.9.** [24, Theorem 4.2] Let  $\mathbf{C}$  be a category that is both monoidal closed and a symmetric monoidal *Met-category*, and suppose that for all objects  $X \in \mathbf{C}$ , the functor  $X \multimap - : \mathbf{C} \rightarrow \mathbf{C}$  is non-expansive. Then  $\mathbf{C}$  is a symmetric monoidal closed *Met-category*.



**Theorem 2.4.10** (Soundness). *[24, Theorem 3.14] The rules in Figures 2 and 3 are sound for a symmetric monoidal closed Met-category  $\mathcal{C}$  over metric spaces. Specifically, if  $\Gamma \triangleright v =_q w : \mathbb{A}$  results from the rules in Figures 2 and 3 then  $q \geq d(\llbracket \Gamma \triangleright v : \mathbb{A} \rrbracket, \llbracket \Gamma \triangleright w : \mathbb{A} \rrbracket)$ .*

**Definition 2.4.11.** Consider a metric  $\lambda$ -theory  $((G, \Sigma), Ax)$  and a symmetric monoidal closed Met-category  $\mathcal{C}$ . Suppose that for each  $X \in G$  we have an interpretation  $\llbracket X \rrbracket$  as a  $\mathcal{C}$ -object and analogously for the operation symbols. This interpretation structure is a model of the theory if all axioms in  $Ax$  are satisfied by the interpretation.

For two types  $\mathbb{A}$  and  $\mathbb{B}$  of a metric  $\lambda$ -theory  $\mathcal{T}$ , consider the class  $\text{Values}(\mathbb{A}, \mathbb{B})$  of values  $v$  such that  $x : \mathbb{A} \triangleright v : \mathbb{B}$ . We equip  $\text{Values}(\mathbb{A}, \mathbb{B})$  with the function  $d : \text{Values}(\mathbb{A}, \mathbb{B}) \times \text{Values}(\mathbb{A}, \mathbb{B}) \rightarrow \mathbb{Q}_0^+$  defined by,

$$d(v, w) = \inf \{q \mid v =_q w \text{ is a theorem of } \mathcal{T}\}$$

$\text{Values}((\mathbb{A}, \mathbb{B}), d)$  is a (possibly large) Met-category. We quotient this Met-category into a separated Met-category which we denote by  $(\text{Values}(\mathbb{A}, \mathbb{B}), d)/\sim$ . We call  $\mathcal{T}$  *varietal* if  $(\text{Values}(\mathbb{A}, \mathbb{B}), d)/\sim$  is a small Met-category for all types  $\mathbb{A}$  and  $\mathbb{B}$ . For the rest of this work, we will focus exclusively on varietal theories and locally small categories.



## Chapter 3

# Metric $\lambda$ -calculus with conditionals

As noted in [subsection 2.2.9](#), a metric equation for the conditional statement would be extremely helpful for reasoning about approximate equivalence within the  $\lambda$ -calculus with conditionals. In this chapter, we address this gap by introducing such an equation and proving both its soundness and completeness. We present the category of metric spaces as an example of a setting suitable for reasoning about approximate equivalence in this context and illustrate the usefulness of the introduced equation through two small examples: reasoning about approximate equivalence between boolean terms (i.e., terms of type  $\mathbb{I} \oplus \mathbb{I}$ ) and the extensionality of copairing. We conclude the chapter with a brief discussion on the suitability of the chosen metric equation.

Discussao sobre  $\max(q+r, q+s) \stackrel{!}{=} \max(q+r, q+s)$

### 3.1 Syntax

The metric equation for conditionals is presented in [Figure 5](#).

$$\frac{v =_q v' \quad w =_r w' \quad u =_s u'}{\text{case } v \{ \text{inl}(x) \Rightarrow w; \text{inr}(y) \Rightarrow u \} =_{q+\sup\{r,s\}} \text{case } v' \{ \text{inl}(x) \Rightarrow w'; \text{inr}(y) \Rightarrow u' \}}$$

Figure 5: Metric equation for condicionals

Intuitively, this equation states that the maximum distance between the two conditional statements is determined the:

- The larger of the maximum distances between the corresponding branches, i.e., the program pairs  $(w, w')$  and  $(u, u')$ ; and

- The maximum distance between the terms  $v$  and  $v'$  on which their execution depends.

Recall [Definition 2.2.20](#). Now, we extend  $Th(Ax)$  to denote the smallest class that contains  $Ax$  and is closed under the rules presented in [Figure 2](#), [Figure 3](#) and [Figure 5](#).

## 3.2 Interpretation

In this subsection, we extend the concepts introduced in ?? to include the interpretation of the new metric equation.

**Definition 3.2.1.** A *distributive symmetric monoidal closed Met-category*  $\mathcal{C}$  is a category that is both distributive symmetric monoidal closed and a symmetric monoidal closed Met-category, such that, for all morphisms  $f_1, f_2 \in \mathcal{C}(A, C)$  and  $g_1, g_2 \in \mathcal{C}(B, C)$ , we have:

$$d_{A \oplus B, C}([f_1, g_1], [f_2, g_2]) \leq \sup\{d_{A, C}(f_1, f_2), d_{B, C}(g_1, g_2)\}.$$

**Definition 3.2.2.** Consider a metric  $\lambda$ -theory  $((G, \Sigma), Ax)$  and a distributive symmetric monoidal closed Met-category  $\mathcal{C}$ . Suppose that for each  $X \in G$  we have an interpretation  $\llbracket X \rrbracket$  as a  $\mathcal{C}$ -object and analogously for the operation symbols. This interpretation structure is a model of the theory if all axioms in  $Ax$  are satisfied by the interpretation.

Não sei será melhor colocar isto elsewhere

## 3.3 Lattice Theory Preliminaries

In this section, we introduce a few concepts from lattice theory that will be useful for the completeness proof and for the broader discussion of our results.

**Definition 3.3.1.** A *lattice* is a partial order in which every finite subset has both a meet and a join. A *complete lattice* is a partial order in which every subset, finite or infinite, has a meet and a join.

**Definition 3.3.2.** *Quantales* are complete lattices equipped with an associative composition, typically denoted  $\otimes$ , that preserves suprema in both arguments.

**Example 3.3.3.** In this work, we consider only the standard metric quantale, whose complete lattice is the interval  $[0, \infty]$  ordered by  $\geq$ . The associative composition is given by addition.

**Lemma 3.3.4** (Way-below relation for the metric quantale). [24] Let  $L$  denote the complete lattice  $[0, \infty]$  ordered by  $\geq$ . For all  $x, y \in L$  and for every subset  $X \subseteq L$ , if  $y > x$ , then whenever  $x \geq \inf X$ , there exists a finite subset  $A \subseteq X$  such that  $y \geq \inf A$ .

This property characterizes the way-below relation in the metric quantale. The lattice  $L$  is continuous, i.e. for every  $x \in L$ ,

$$x = \inf\{y \mid y \in L \text{ and } y > x\}.$$

**Definition 3.3.5.** A subset  $D$  of a lattice  $L$  is called *directed* if it is nonempty and every finite subset of  $D$  has an upper bound in  $D$ . A partially order is said to be *directed complete* if every directed subset has a supremum. A directed complete poset is commonly referred to as a *dcpo*.

**Definition 3.3.6.** A semilattice  $L$  is called *meet continuous* if it is directed complete, i.e., a dcpo, and satisfies the condition

$$x \wedge \sup D = \sup\{x \wedge d \mid d \in D\} \quad (\text{MC})$$

for all  $x \in L$  and all directed subsets  $D \subseteq L$ .

**Lemma 3.3.7.** [41, Proposition I-1.8] Every continuous lattice is meet continuous.

**Lemma 3.3.8.** [42, Lemma 2.23] Let  $P$  be a lattice, let  $S, T \subseteq P$  and assume that  $\bigvee S, \bigvee T, \bigwedge S$  and  $\bigwedge T$  exist in  $P$ . Then

$$\bigvee(S \cup T) = \left(\bigvee S\right) \vee \left(\bigvee T\right) \quad \text{and} \quad \bigwedge(S \cup T) = \left(\bigwedge S\right) \wedge \left(\bigwedge T\right).$$

## 3.4 Soundness and Completeness

The proofs in this section are based on the proofs of Theorem 2.4.10 and Theorem 3.4.2 in [24].

**Theorem 3.4.1.** The rules in Figures 2, 3 and 5 are sound for a distributive symmetric monoidal closed Met-category  $\mathcal{C}$ . Specifically, if  $\Gamma \triangleright v =_q w : \mathbb{A}$  results from the rules in Figures 2, 3 and 5 then  $q \geq d(\llbracket \Gamma \triangleright v : \mathbb{A} \rrbracket, \llbracket \Gamma \triangleright w : \mathbb{A} \rrbracket)$ .

*Proof.* As is shown in [24], this proof uses induction over the depth of proof trees that arise from the metric deductive system. The general strategy for each inference rule is to use the definition of a distributive symmetric monoidal closed Met-category.

More concretely, first, consider the equations on [Figure 2](#) which abbreviate equations  $\Gamma \triangleright w =_0 v : \mathbb{A}$ . By [Theorem 2.4.2](#), these equations are sound for distributive symmetric monoidal closed categories, i.e. if  $v = w$ , then  $\llbracket v \rrbracket = \llbracket w \rrbracket$  in  $\mathbf{C}$ . Then by the definition of a Met-Category ([Definition 2.4.8](#)) we obtain  $d(\llbracket v \rrbracket, \llbracket w \rrbracket) = d(\llbracket w \rrbracket, \llbracket v \rrbracket) = 0$ .

The rules in [Figure 3](#) follows from the definition of a symmetric monoidal closed Met-category. Finally, regarding the rule in [Figure 5](#), it follows from the fact that  $\mathbf{C}$  is distributive symmetric monoidal closed Met-category. We reason as follows:

$$\begin{aligned}
& d(\llbracket \text{case } v \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow w; \text{inr}_{\mathbb{A}}(y) \Rightarrow u \} \rrbracket, \llbracket \text{case } v' \{ \text{inl}_{\mathbb{B}}(x) \Rightarrow w'; \text{inr}_{\mathbb{A}}(y) \Rightarrow u' \} \rrbracket) \\
&= d(\llbracket w \rrbracket, \llbracket u \rrbracket \cdot (\text{jn}_{\Delta; \mathbb{A}} \cdot \text{sw} \oplus \text{jn}_{\Delta; \mathbb{B}} \cdot \text{sw}) \cdot \text{dist} \cdot (\llbracket v \rrbracket \otimes \text{id}) \cdot \text{sp}_{\Gamma; \Delta} \cdot \text{sh}_E, \\
&\quad \llbracket w' \rrbracket, \llbracket u' \rrbracket \cdot (\text{jn}_{\Delta; \mathbb{A}} \cdot \text{sw} \oplus \text{jn}_{\Delta; \mathbb{B}} \cdot \text{sw}) \cdot \text{dist} \cdot (\llbracket v' \rrbracket \otimes \text{id}) \cdot \text{sp}_{\Gamma; \Delta} \cdot \text{sh}_E) \\
&\leq d(\llbracket w \rrbracket, \llbracket u \rrbracket \cdot (\text{jn}_{\Delta; \mathbb{A}} \cdot \text{sw} \oplus \text{jn}_{\Delta; \mathbb{B}} \cdot \text{sw}) \cdot \text{dist} \cdot (\llbracket v \rrbracket \otimes \text{id}), \llbracket w' \rrbracket, \llbracket u' \rrbracket \cdot (\text{jn}_{\Delta; \mathbb{A}} \cdot \text{sw} \oplus \\
&\quad \text{jn}_{\Delta; \mathbb{B}} \cdot \text{sw}) \cdot \text{dist} \cdot (\llbracket v' \rrbracket \otimes \text{id})) \\
&\leq d(\llbracket v \rrbracket \otimes \text{id}, \llbracket v' \rrbracket \otimes \text{id}) + d(\llbracket w \rrbracket, \llbracket u \rrbracket \cdot (\text{jn}_{\Delta; \mathbb{A}} \cdot \text{sw} \oplus \text{jn}_{\Delta; \mathbb{B}} \cdot \text{sw}) \cdot \text{dist}, \llbracket w' \rrbracket, \llbracket u' \rrbracket \cdot \\
&\quad (\text{jn}_{\Delta; \mathbb{A}} \cdot \text{sw} \oplus \text{jn}_{\Delta; \mathbb{B}} \cdot \text{sw}) \cdot \text{dist}) \\
&\leq q + d(\llbracket w \rrbracket, \llbracket u \rrbracket \cdot (\text{jn}_{\Delta; \mathbb{A}} \cdot \text{sw} \oplus \text{jn}_{\Delta; \mathbb{B}} \cdot \text{sw}) \cdot \text{dist}, \llbracket w' \rrbracket, \llbracket u' \rrbracket \cdot (\text{jn}_{\Delta; \mathbb{A}} \cdot \text{sw} \oplus \text{jn}_{\Delta; \mathbb{B}} \cdot \text{sw}) \\
&\quad \cdot \text{dist}) \\
&\leq q + d(\llbracket w \rrbracket, \llbracket u \rrbracket, \llbracket w' \rrbracket, \llbracket u' \rrbracket) \\
&\leq q + \sup(d(\llbracket w \rrbracket, \llbracket w' \rrbracket), d(\llbracket u \rrbracket, \llbracket u' \rrbracket)) \\
&\leq q + \sup(r, s)
\end{aligned}$$

The second step follows from the fact that  $\text{sp}_{\Gamma; \Delta} \cdot \text{sh}_E$  is a morphism in  $\mathbf{C}$  and that  $\mathbf{C}$  is a Met-category. The third and fifth step follow from an analogous reasoning. The fourth step follows from the premises of the rule in question and the fact that  $\mathbf{C}$  is a symmetric monoidal Met-category. The sixth step follows from the fact that  $\mathbf{C}$  is a distributive symmetric monoidal Met-enriched. Finally, the last step follows from the premise of the rule in question. □

Here, we note that in  $\text{Values}((\mathbb{A}, \mathbb{B}), d)$ , the definition of  $d$  is extended incorporating the new theorems and we quotient this Met-category into a separated Met-category which we denote by  $(\text{Values}(\mathbb{A}, \mathbb{B}), d)/\sim$ .

**Theorem 3.4.2** (Completeness). [24, Theorem 3.16] Consider a varietal metric  $\lambda$ -theory. A metric equation in context  $\Gamma \triangleright v =_q w : \mathbb{A}$  is a theorem if and only if it holds in all models of the theory.

*Proof.* Completeness arises from constructing the syntactic category  $\text{Syn}(\mathcal{T})$  of  $\lambda$ -theory  $\mathcal{T}$  of the underlying theory  $\mathcal{T}$  and then show that provability of  $\Gamma \triangleright v =_q w : \mathbb{A}$  in  $\mathcal{T}$  is equivalent to  $a(\llbracket v \rrbracket, \llbracket w \rrbracket) \geq q$  in the category  $\text{Syn}(\mathcal{T})$ . Here we take advantage of the quotienting of a Met-category into a separated Met-category to prove that. First consider the following diagram:

$$\begin{array}{ccc} X \times X & \xrightarrow{\quad} & X/\sim \times X/\sim \\ \downarrow f & \nearrow \hat{f} & \\ Y & & \end{array}$$

It is necessary to prove that when  $\hat{f}$  is defined as  $\hat{f}([x], [y]) = [f(x, y)]$ ,  $\hat{f}$  is a well-defined function. This is the case because if  $[x] = [x']$  and  $[y] = [y']$ , then  $x =_0 x'$  and  $y =_0 y'$ , and consequently applying rule (refl) we obtain that  $f(x, y) =_0 f(x', y')$ . Therefore, the quotienting process identifies all terms  $x : \mathbb{A} \triangleright v : \mathbb{B}$  and  $x : \mathbb{A} \triangleright w : \mathbb{B}$  such that  $v =_0 w$  and  $w =_0 v$ . This relation includes the equations-in-context from ?? and, attending to Theorem 2.4.4 it is compatible with the term formation rules of the linear  $\lambda$ -calculus ?. Thus, we obtain a category  $\text{Syn}(\mathcal{T})$  whose objects are the types of the language and whose hom-sets are the underlying sets of the Met-categories  $(\mathbf{Values}(\mathbb{A}, \mathbb{B}), d)/\sim$ .

The next step is to show that  $\text{Syn}(\mathcal{T})$  is a distributive symmetric monoidal Met-category. To this effect, in general one uses the fact that  $+$  distributes over suprema and that  $q \wedge (-)$  distributes over suprema for every  $q \in [0, \infty]$  (Lemma 3.3.7). With respect to the case statement, we reason as follows:

$$\begin{aligned} & \sup \{d([v], [w]), d([v'], [w'])\} \\ &= \sup \{d(v, w), d(v', w')\} \\ &= \sup \{\inf \{q \mid v =_q v'\}, \inf \{r \mid w =_r w'\}\} \\ &= \inf \{\sup \{q, r\} \mid v =_q v', w =_r w'\} \\ &\geq \inf \{q \mid \text{case } z \{\text{inl}(x) \Rightarrow v; \text{inr}(y) \Rightarrow w\} =_q \text{case } z \{\text{inl}(x) \Rightarrow v'; \text{inr}(y) \Rightarrow w'\}\} \end{aligned}$$

$$\begin{aligned}
&= d(\text{case } z \{ \text{inl}(x) \Rightarrow v; \text{inr}(y) \Rightarrow w \}, \text{case } z \{ \text{inl}(x) \Rightarrow v'; \text{inr}(y) \Rightarrow w' \}) \\
&= d([\text{case } z \{ \text{inl}(x) \Rightarrow v; \text{inr}(y) \Rightarrow w \}], [\text{case } z \{ \text{inl}(x) \Rightarrow v'; \text{inr}(y) \Rightarrow w' \}]) \\
&= d([v], [v']), [[w], [w']]
\end{aligned}$$

The third step follows from the fact that  $q \wedge (-)$  distributes over suprema for every  $q \in [0, \infty]$ , and the fourth step follows from the fact that for any sets  $A$  and  $B$ , if  $A \subseteq B$ , then  $\inf\{A\} \geq \inf\{B\}$ .

The final step is to show that if an equation  $\Gamma \triangleright v =_q v' : \mathbb{A}$  with  $q \in \mathbb{Q}_0^+$  is satisfied by  $\text{Syn}(\mathcal{T})$  then it is a theorem of the linear metric  $\lambda$ -theory. By assumption,  $d([v], [v']) = d(v, v') = \inf \{r \mid v =_r v'\} \leq q$ . It follows from the definition of the way-below relation that for all  $x \in \mathbb{Q}_0^+$  with  $x > q$  there exists a finite set  $A \subseteq \{r \mid v =_r v'\}$  such that  $x \geq \inf A$ . Then by an application of rule (join) we obtain  $v =_{\inf A} v'$ , and consequently, rule (weak) provides  $v =_x v'$  for all  $x > q$ . Finally, by an application of rule (arch) we deduce that  $v =_q v'$  is part of the theory.  $\square$

### 3.5 A small example: Met

**Proposition 3.5.1.** *The category Met is a distributive symmetric monoidal closed Met-category*

*Proof.* By [24, Example 3.8], the category Met is a symmetric monoidal closed Met-category. As a result, it suffices to show that for all morphisms  $f_1, f_2 \in \text{Met}(A, C)$  and  $g_1, g_2 \in \text{Met}(B, C)$ , the following inequality holds:

$$d_{A \oplus B, C}([f_1, g_1], [f_2, g_2]) \leq \sup\{d_{A, C}(f_1, f_2), d_{B, C}(g_1, g_2)\}.$$

In this category, the coproduct is defined as in Set. The distance function  $d$  on the coproduct  $A \oplus B$  is given by:

$$\begin{cases} d_{A \oplus B}(\text{inl}(a_1), \text{inl}(a_2)) = d_A(a_1, a_2) \\ d_{A \oplus B}(\text{inr}(b_1), \text{inr}(b_2)) = d_B(b_1, b_2) \\ d_{A \oplus B}(\text{inl}(a), \text{inr}(b)) = d_{A \oplus B}(\text{inr}(b), \text{inl}(a)) = \infty \end{cases}$$

The co-pairing is defined as in Set. The inequality we aim to prove follows directly from the fact that, given two morphisms  $f, g \in \text{Met}(A, B)$  the distance between them is defined as

$$d_B(\text{inl}(x_1), \text{inl}(x_2)) = \sup\{d_A(fa, ga) \mid a \in A\},$$



together with [Lemma 3.3.8](#). For  $f_1, f_2 \in \text{Met}(A, C)$  and  $g_1, g_2 \in \text{Met}(B, C)$ , we calculate:

$$\begin{aligned}
& d_{A \oplus B, C}([f_1, g_1], [f_2, g_2]) \\
&= \sup\{d_{A \oplus B}([f_1, g_1](a, b), [f_2, g_2](a, b)) \mid (a, b) \in A \oplus B\} \\
&= \sup\{\{d_{A \oplus B}([f_1, g_1](\text{inl}(a)), [f_2, g_2](\text{inl}(a))) \mid a \in A\} \\
&\quad \cup \{d_{A \oplus B}([f_1, g_1](\text{inr}(b)), [f_2, g_2](\text{inr}(b))) \mid b \in B\}\} \\
&= \sup\{\{d_A(f_1(a), f_2(a)) \mid a \in A\} \cup \{d_B(g_1(b), g_2(b)) \mid b \in B\}\} \\
&= \sup\{\sup\{d_A(f_1(a), f_2(a)) \mid a \in A\}, \sup\{d_B(g_1(b), g_2(b)) \mid b \in B\}\} \\
&= \sup\{d_{A, C}(f_1, f_2), d_{B, C}(g_1, g_2)\}.
\end{aligned}$$

□

### 3.6 Interlude:Booleans - Part 2

Having proven that the metric equation for conditionals is both sound and complete, we can now use it to explore the booleans introduced in the previous chapter in [subsection 2.2.9](#).

For instance, given axioms

$$v' =_{\epsilon_1} a \quad \text{and} \quad w' =_{\epsilon_2} b$$

we can postulate

$$\mathbf{Conjunction}[v'/v, w'/w] =_{\epsilon_1 + \epsilon_2} \mathbf{Conjunction}[a/v, b/w].$$

First note that using the equation  $\eta_{\text{dis}}$ , we have that  $\Gamma \triangleright \text{dis}(w) =_0 \text{dis}(b)$ . Then, using our metric deductive system we calculate

$$\begin{aligned}
& \text{case } v' \left\{ \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . w; \text{inl}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . \text{dis}(w) \text{ to } * . \text{inr}_{\mathbb{I}}(*) \right\} \\
&=_{\epsilon_1 + \epsilon_2} \text{case } a \left\{ \text{inl}_{\mathbb{I}}(x) \Rightarrow x \text{ to } * . b; \text{inl}_{\mathbb{I}}(y) \Rightarrow y \text{ to } * . \text{dis}(b) \text{ to } * . \text{inr}_{\mathbb{I}}(*) \right\}.
\end{aligned}$$

Applying similar reasoning, is straightforward to see that we have a similar equation for the

**Disjunction** program and that given an axiom  $v' =_{\epsilon} a$ , we can postulate **Negation** $[v'/v] =_{\epsilon}$  **Negation** $[a/v]$ .

Now, consider the axiom

$$v' =_{\epsilon} \text{inl}_{\mathbb{I}}(*)$$

Considering Lemma 2.2.14, we have

$$\mathbf{Conjunction}[v'/v, w'/w] =_{\epsilon} \mathbf{Conjunction}[\text{inl}_{\mathbb{I}}(*)'/v, w'/w] = w'$$

### 3.7 Extensionality of the copairing

A  $\lambda$ -abstraction that receives inputs of a disjunctive type is determined by what it does to inputs “from the left and from the right”, *i.e.*,

$$\begin{cases} (\lambda x.t) \text{inl}(y) = (\lambda x.s) \text{inl}(y) \\ (\lambda x.t) \text{inr}(z) = (\lambda x.s) \text{inr}(z) \end{cases} \implies \lambda x.t = \lambda x.s$$

The proof is as follows:

Using the  $\beta$  equation we have

$$\begin{cases} (\lambda x.t) \text{inl}(y) = (\lambda x.s) \text{inl}(y) \\ (\lambda x.t) \text{inr}(z) = (\lambda x.s) \text{inr}(z) \end{cases} = \begin{cases} t [\text{inl}(y)/x] = s [\text{inl}(y)/x] \\ t [\text{inr}(z)/x] = s [\text{inr}(z)/x] \end{cases} \quad (3.1)$$

Next, considering the equations above,  $\eta_{case}$ , and the  $\alpha$ -equivalence we reason as follows:

$$\begin{aligned} t &= t[x'/x] = \text{case } x' \left\{ \text{inl}_{\mathbb{I}}(y) \Rightarrow t [\text{inl}(y)/x]; \text{inl}_{\mathbb{I}}(z) \Rightarrow t [\text{inr}(z)/x] \right\} & (\alpha, \eta_{case}) \\ &= \text{case } x' \left\{ \text{inl}_{\mathbb{I}}(y) \Rightarrow s [\text{inl}(y)/x]; \text{inl}_{\mathbb{I}}(z) \Rightarrow s [\text{inr}(z)/x] \right\} = s & (\text{Equation 3.1}, \eta_{case}, \alpha) \end{aligned}$$

Finally, it is straightforward that  $t = s \implies \lambda x.t = \lambda x.s$ .

Now assume that

$$\begin{cases} (\lambda x.t) \text{inl}(y) =_{\epsilon_1} (\lambda x.s) \text{inl}(y) \\ (\lambda x.t) \text{inr}(z) =_{\epsilon_2} (\lambda x.s) \text{inr}(z) \end{cases} \implies \lambda x.t = \lambda x.s$$

Following the same reasoning as before and applying the metric equation for conditionals we obtain:

$$\begin{aligned} t &= \text{case } x' \left\{ \text{inl}_{\mathbb{I}}(y) \Rightarrow t [\text{inl}(y)/x]; \text{inl}_{\mathbb{I}}(z) \Rightarrow t [\text{inr}(z)/x] \right\} & (\alpha, \eta_{case}) \\ &=_{\max\{\epsilon_1, \epsilon_2\}} \text{case } x' \left\{ \text{inl}_{\mathbb{I}}(y) \Rightarrow s [\text{inl}(y)/x]; \text{inl}_{\mathbb{I}}(z) \Rightarrow s [\text{inr}(z)/x] \right\} = s & (\text{Figure 5}, \eta_{case}, \alpha) \end{aligned}$$

As a result, given our metric deductive system we obtain  $\lambda x.t =_{\max\{\epsilon_1, \epsilon_2\}} \lambda x.s$ .

### 3.8 Some remarks on the metric equation chosen

Falta exemplo de quantale em que não yemos distributividade

One might wonder why, in the definition of the metric, we chose to use the expression  $q + \sup\{r, s\}$  rather than  $\sup\{q + r, q + s\}$ , given that these two expressions are, in fact, equal in the metric quantale. More generally, this equality corresponds to the identity

$$q \otimes \inf\{r, s\} = \inf\{q \otimes r, q \otimes s\},$$

for  $r, s \in L$ , where  $L$  denotes the underlying lattice of the quantale. (Note again that in the case of the metric quantale, the usual order is reversed, so infima correspond to suprema in the standard order.)

However, this identity does *not* hold in arbitrary quantales **[CONTRAEXAMPLE NEEDED]**. Moreover, as becomes clear in the soundness proof, it is the expression  $q \otimes \sup\{r, s\}$  that arises naturally.



# **Part II**

## **Applications**



## Chapter 4

# Probabilistic Programming

Probabilistic programs offer a structured approach to drawing statistical conclusions from uncertain data or real-world observations. They generalize probabilistic graphical models beyond the capabilities of Bayesian networks and are expected to have broader applications in machine intelligence. These programs are employed across various domains. They control autonomous systems, verify security protocols, and implement randomized algorithms for solving computationally intractable problems. As a result, they are becoming increasingly central to AI development. At their core, they aim to democratize probabilistic modeling by providing programmers with expressive, high-level abstractions for machine learning and statistical reasoning [43].

The next section introduces the fundamentals of measure theory, a key component in defining the semantics of probabilistic programs, drawing inspiration primarily from [44] and [45]. For a more in-depth exploration of measure theory, see reference [46].

Meter esta intro direita

## 4.1 Banach spaces

### 4.1.1 Normed spaces

The definition of normed spaces given in Definition 5.1.10 extends to infinite-dimensional vector spaces.

**Definition 4.1.1.** Let  $V$  and  $W$  be normed vector spaces, and let  $T : V \rightarrow W$  be a linear operator. The *operator norm*, denoted by  $\|\cdot\|_{\text{op}}$ , is defined as

$$\|T\|_{\text{op}} = \sup\{\|T(v)\| : v \in V, \|v\| = 1\}.$$

If  $\|T\|_{\text{op}} < \infty$ , we say that  $T$  is a *bounded operator*; otherwise, if  $\|T\|_{\text{op}} = \infty$ , we say that  $T$  is *unbounded*. We denote by  $\mathcal{B}(V, W)$  the vector space of all bounded linear operators from  $V$  to  $W$ , and we write  $\mathcal{B}(V)$  for  $\mathcal{B}(V, V)$ .

**Lemma 4.1.2.** *Let  $T : V \rightarrow W$  be a bounded linear operator between normed spaces. Then the following statements hold:*

1. *For every  $v \in V$ , we have  $\|T(v)\| \leq \|T\|_{\text{op}} \cdot \|v\|$ .*
2. *The operator  $T$  is continuous if and only if it is bounded.*

**Definition 4.1.3.** Let  $V$  and  $W$  be normed vector spaces, and let  $T : V \rightarrow W$  be a linear operator.  $T$  is called a *short map* if  $\|T\|_{\text{op}} \leq 1$ .

### 4.1.2 Banach spaces

**Definition 4.1.4.** (*Cauchy sequence*) Suppose  $d$  is a metric on a set  $X$ . A sequence  $\{x_n\} \subset X$  is called a *Cauchy sequence* if, for every  $\varepsilon > 0$ , there exists an integer  $N \in \mathbb{N}$  such that  $d(x_m, x_n) < \varepsilon$  for all  $m, n > N$ . The metric  $d$  is said to be *complete* if every Cauchy sequence in  $X$  converges to a point in  $X$ .

**Definition 4.1.5.** A *Banach space* is a normed vector space that is complete with respect to the metric induced by its norm. In other words, every Cauchy sequence in the space must converge to a limit within the space.

**Definition 4.1.6** (Algebraic Tensor Product). Let  $V_1, \dots, V_n$  and  $W$  be vector spaces, and let

$$\varphi : V_1 \times \cdots \times V_n \rightarrow W$$

be a multilinear function, meaning a function for which the mapping

$$u_k \mapsto \varphi(u_1, \dots, u_n)$$

is linear for each  $k \in \{1, \dots, n\}$  and every fixed choice of vectors  $u_1, \dots, u_{k-1}, u_{k+1}, \dots, u_n$ . Then there exists a unique linear mapping

$$A : V_1 \otimes \cdots \otimes V_n \rightarrow W$$

such that

$$\varphi(u_1, \dots, u_n) = A(u_1 \otimes \cdots \otimes u_n)$$

for all choices of  $u_1 \in V_1, \dots, u_n \in V_n$ .



**Definition 4.1.7.** Let  $V$  and  $W$  be Banach spaces. Let  $u$  be any element of  $V \otimes W$ . The *projective norm*, denoted  $\|\cdot\|_\pi$ , is defined by:

$$\|u\|_\pi = \inf \left\{ \sum_{i=1}^n \|v_i\| \|w_i\| \mid u = \sum_{i=1}^n v_i \otimes w_i \right\}.$$

**Proposition 4.1.8.** Let  $V$  and  $W$  be Banach spaces. Then, for every  $v \in V$  and  $w \in W$ , it holds that  $\|v \otimes w\|_\pi = \|v\| \otimes \|w\|$ .

**Definition 4.1.9.** Let  $V$  and  $W$  be Banach spaces. The *projective tensor product* of  $V$  and  $W$ , denoted  $V \widehat{\otimes}_\pi W$ , is the completion of the algebraic tensor product  $V \otimes W$  with respect to the projective norm  $\|\cdot\|_\pi$ .

## 4.2 Category Ban

**Example 4.2.1.** The category Ban is the category of Banach spaces and short maps.

**Example 4.2.2.** In the category Ban, the coproduct of Banach spaces is given by their direct sum equipped with the  $L_1$ -norm. The injections and copairing map are defined analogously to those in CVect.

## 4.3 Measure theory

Probabilistic computation involves running programs incorporating randomness, leading to output behaviors characterized by probability distributions rather than deterministic outcomes. To effectively understand and analyze these programs, it is essential to have a solid foundation in reasoning about probability distributions. That is where measure theory comes into play.

Acrescentar medidas negativas tb

### 4.3.1 What is measure theory?

Throughout history, mathematicians sought to extend the ideas of length, area, and volume. The most effective way to generalize these concepts is through the idea of a measure. Abstractly, a measure is a function defined on sets with additive properties mirroring length, area, and volume.

We begin with a simple example inspired by [45] to develop an intuition for the concepts of measure and measure space. Imagine an open field  $S$  covered in snow after a storm. Suppose we wish to measure the amount of snow accumulated in as many field regions as possible. Assume we have accurate tools for measuring snow over standard geometric shapes like triangles, rectangles, and circles. We can approximate irregularly shaped regions using combinations of these standard shapes and then apply a limiting process to assign a consistent measure to such regions. Let  $\mathcal{B}$  denote the collection of subsets of  $S$  that are deemed *measurable*, let  $\lambda(A)$  represent the amount of snow in each  $A \in \mathcal{B}$ , and let  $A^c$  denote the complement of a set  $A$ .

For this framework to make sense, it is reasonable to require that  $\mathcal{B}$  and  $\lambda(\cdot)$  satisfy the following properties:

#### Properties of $\mathcal{B}$ :

1. If  $A \in \mathcal{B}$ , then the complement  $A^c \in \mathcal{B}$ . (i.e., if we can measure the snow on a set  $A$ , and we know the total amount on  $S$ , then we can determine the snow on the remaining part  $A^c$ .)
2. If  $A_1, A_2 \in \mathcal{B}$ , then  $A_1 \cup A_2 \in \mathcal{B}$ . (i.e., if we can measure the snow on two regions  $A_1$  and  $A_2$ , we should also be able to measure it on their union.)
3. If  $\{A_n\}_{n \geq 1} \subset \mathcal{B}$  is an increasing sequence, i.e.,  $A_n \subset A_{n+1}$  for all  $n$ , then  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$ . (i.e., if each set in a increasing sequence of regions is measurable, then their limit—the union—should also be measurable.)
4. The collection  $\mathcal{B}$  contains a base class  $C$  of simple, well-behaved sets (e.g., triangles, rectangles, circles) for which measurement is initially defined.

#### Properties of $\lambda(\cdot)$ :

1.  $\lambda(A) \geq 0$  for all  $A \in \mathcal{B}$  (i.e., the amount of snow on any set must be nonnegative).
2. If  $A_1, A_2 \in \mathcal{B}$  and  $A_1 \cap A_2 = \emptyset$ , then  $\lambda(A_1 \cup A_2) = \lambda(A_1) + \lambda(A_2)$  (i.e., The total amount of snow over two non-overlapping regions is just the sum of the snow in each region. This characteristic of  $\lambda$  is known as *finite additivity*.)

3. If  $\{A_n\}_{n \geq 1} \subset \mathcal{B}$  is an increasing sequence, i.e.,  $A_n \subset A_{n+1}$  then  $\lambda(\lim_{n \rightarrow \infty} A_n) = \lambda(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \lambda(A_n)$ . (i.e., if a set can be approximated by an increasing sequence of measurable sets  $\{A_n\}_{n \geq 1}$ , then  $\lambda(A) = \lim_{n \rightarrow \infty} \lambda(A_n)$ . This is known as *monotone continuity from below*.)

### 4.3.2 Measurable spaces and measures

Remarkably, these intuitive conditions give rise to a profoundly versatile and far-reaching theoretical framework. The requirements imposed on  $\mathcal{B}$  and  $\lambda$  can equivalently be stated as follows:

#### Properties of $\mathcal{B}$ :

1.  $\emptyset \in \mathcal{B}$ .
2.  $A \in \mathcal{B} \implies A^c \in \mathcal{B}$ .
3.  $A_1, A_2, \dots \in \mathcal{B} \implies \bigcup_i A_i \in \mathcal{B}$  (this is known as *closure under countable unions*).

#### Properties of $\lambda$ :

1.  $\lambda(\cdot) \geq 0$  and  $\lambda(\emptyset) = 0$ .
2. If  $\{A_n\}_{n \geq 1} \subset \mathcal{B}$  is a sequence of pairwise disjoint sets (i.e.,  $A_i \cap A_j = \emptyset$  for  $i \neq j$ ), then  $\lambda(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \lambda(A_n)$  (this is known as *countable additivity*).

A collection  $\mathcal{B}$  of subsets of  $S$  satisfying the above conditions for  $\mathcal{B}$  is designated a  $\sigma$ -algebra. Similarly, a set function  $\lambda$  defined on a  $\sigma$ -algebra  $\mathcal{B}$  that fulfills the above properties for  $\lambda$  qualifies as a *measure*.

**Definition 4.3.1.** A  $\sigma$ -algebra  $\mathcal{B}$  on a set  $S$  is a collection of subsets of  $S$  that includes the empty set, is closed under complementation with respect to  $S$ , and is closed under countable unions.

**Definition 4.3.2.** The pair  $(S, \mathcal{B})$  where  $\mathcal{B}$  is a  $\sigma$ -algebra constitutes a *measurable space*. The elements of  $\mathcal{B}$  are called the *measurable sets* of the space.

**Definition 4.3.3.** The Borel  $\sigma$ -algebra of  $\mathbb{R}^n$  is the  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R}^n)$  generated by all open sets. The sets in  $\mathcal{B}(\mathbb{R}^n)$  are called Borel sets.

**Definition 4.3.4.** A *measure* on the measurable space  $(S, \mathcal{B})$  is a function  $\mu : \mathcal{B} \rightarrow [0, \infty]$  that is countably additive and satisfies  $\mu(\emptyset) = 0$ . A measure on  $(S, \mathcal{B})$  is called a *probability measure* if it assigns total measure 1 to the entire space, that is,  $\mu(S) = 1$ .

In the context of probability theory, such measures are often referred to as *distributions*. In what follows, we will use the terms measure and distribution interchangeably.

One of the most important measures is the Lebesgue measure on the real line (i.e., the length in  $\mathbb{R}$ ), and its generalizations to  $\mathbb{R}^n$ . It is characterized as the unique measure on the Borel sets, whose value on every interval is its length. That is, for any interval  $(a, b) \subset \mathbb{R}$ ,  $\lambda((a, b)) = b - a$ .

The Lebesgue measure is *translation-invariant*, meaning that shifting a set does not change its measure, and it is also  *$\sigma$ -finite*.

For any  $s \in S$ , the *Dirac measure* (also known as the *Dirac delta* or *point mass* at  $s$ ) is the probability measure defined by

$$\delta_s(B) = \begin{cases} 1, & \text{if } s \in B, \\ 0, & \text{if } s \notin B. \end{cases}$$

A measure is called *discrete* if represented as a countable weighted sum of Dirac measures. In particular, a *convex combination* of Dirac measures yields a discrete probability measure. These are of the form  $\sum_{s \in C} a_s \delta_s$ , where  $a_s \geq 0$  for all  $s \in C$ , and the weights satisfy  $\sum_{s \in C} a_s = 1$ .

On the other hand, a measure  $\mu$  on a measurable space  $(S, \mathcal{B})$  is called *continuous* if it assigns zero measure to all singleton sets, i.e.,  $\mu(\{s\}) = 0$  for every  $s \in S$ . An example of a continuous measure is the *Lebesgue measure* on  $\mathbb{R}^n$  (for  $n \in \mathbb{N}$ ).

**Definition 4.3.5.** Let  $(S, \mathcal{B}_S)$  and  $(T, \mathcal{B}_T)$  be measurable spaces. Given a function  $f : (S, \mathcal{B}_S) \rightarrow (T, \mathcal{B}_T)$  and a measure  $\mu$  on  $\mathcal{B}_S$ , the *pushforward measure*  $f_*(\mu)$  on  $\mathcal{B}_T$  is defined by:

$$f_*(\mu)(B) = \mu(f^{-1}(B)), \quad B \in \mathcal{B}_T.$$

**Definition 4.3.6.** Let  $(S, \mathcal{B}_S)$  and  $(T, \mathcal{B}_T)$  be measurable spaces. A function  $f : S \rightarrow T$  is said to be *measurable* if for every measurable subset  $B \in \mathcal{B}_T$ , the preimage  $f^{-1}(B) \in \mathcal{B}_S$ .

**Definition 4.3.7.** The *Lebesgue integral* generalizes the familiar Riemann integral. Consider a measurable space  $(S, \mathcal{B})$  and a bounded measurable function  $f : S \rightarrow \mathbb{R}_0^+$ , with upper and

lower bounds  $M$  and  $m$ , respectively. The *Lebesgue integral* of  $f$  with respect to a measure  $\mu: \mathcal{B} \rightarrow \mathbb{R}_0^+$ , denoted  $\int f d\mu$ , is defined as the limit of finite weighted sums of the form:

$$\sum_{i=0}^n f(s_i) \mu(B_i),$$

where  $\{B_0, \dots, B_n\}$  forms a measurable partition of  $S$ , and within each  $B_i$ , the variation of  $f$  does not exceed  $(M - m)/n$ . Here,  $s_i \in B_i$  for each  $i$ , and the limit is taken over increasingly refined partitions.

In the case of a finite discrete space  $n = \{1, 2, \dots, n\}$ , the Lebesgue integral simplifies to a weighted sum:

$$\int f d\mu = \sum_{i=1}^n f(i) \mu(i).$$

Given two measurable spaces  $(S_1, \mathcal{B}_1)$  and  $(S_2, \mathcal{B}_2)$ , their product is the measurable space  $(S_1 \times S_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ , where  $S_1 \times S_2$  is the cartesian product and  $\mathcal{B}_1 \otimes \mathcal{B}_2$  is the  $\sigma$ -algebra generated by all measurable rectangles  $B_1 \times B_2$  with  $B_1 \in \mathcal{B}_1$  and  $B_2 \in \mathcal{B}_2$ :

$$\mathcal{B}_1 \otimes_{\text{meas}} \mathcal{B}_2 := \sigma(\{B_1 \times B_2 \mid B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2\}).$$

Measures on this product space are called *joint distributions*, and are uniquely determined by their values on measurable rectangles due to the inductive structure of the product  $\sigma$ -algebra. *Product measures* are a significant class of joint distributions and are defined from measures

**Definition 4.3.8.** Let  $(S_1, \mathcal{B}_1)$  and  $(S_2, \mathcal{B}_2)$  be measurable spaces, and let  $\mu_1$  and  $\mu_2$  be measures on these spaces, respectively. The *product measure*  $\mu_1 \otimes_{\text{meas}} \mu_2$  is defined on measurable rectangles by

$$(\mu_1 \otimes_{\text{meas}} \mu_2)(B_1 \times B_2) = \mu_1(B_1) \mu_2(B_2).$$

This definition extends uniquely to a joint distribution  $\mu_1 \otimes \mu_2 : \mathcal{B}_1 \otimes \mathcal{B}_2 \rightarrow \mathbb{R}_0^+$ , and reflects the notion of probabilistic independence: sampling from  $\mu_1 \otimes \mu_2$  is equivalent to independently sampling from  $\mu_1$  and  $\mu_2$ .

**Definition 4.3.9.** Let  $(S, \mathcal{B}_S)$  and  $(T, \mathcal{B}_T)$  be measurable spaces. A *Markov kernel* is a mapping  $P : S \times \mathcal{B}_T \rightarrow [0, 1]$  satisfying the following two properties:

1. For each fixed  $s \in S$ , the function  $P(s, \cdot) : \mathcal{B}_T \rightarrow [0, 1]$  is a probability measure on  $(T, \mathcal{B}_T)$ .

2. For each fixed  $A \in \mathcal{B}_T$ , the function  $P(\cdot, A) : S \rightarrow [0, 1]$  is a measurable function on  $(S, \mathcal{B}_S)$ .

By a slight abuse of terminology, we will also use the term *Markov kernel* to refer to a mapping  $P' : S \rightarrow (\mathcal{B}_T \rightarrow [0, 1])$   $P'(s) = \mu$ , when  $P : S \times \mathcal{B}_T \rightarrow [0, 1]$ ,  $P(s, A) = \mu(A)$  is a Markov kernel in the definition above.

By a slight abuse of terminology, we will also refer to a mapping  $P' : S \rightarrow (\mathcal{B}_T \rightarrow [0, 1])$   $P'(s) = \mu$ , as a Markov kernel, when there exists a Markov kernel  $P : S \times \mathcal{B}_T \rightarrow [0, 1]$   $P(s, A) = \mu(A)$  for all  $s \in S$  and  $A \in \mathcal{B}_T$ .

**Definition 4.3.10.** Given a measure  $\mu$  on  $S$  and a Markov kernel  $P : S \times \mathcal{B}_T \rightarrow [0, 1]$ , we can define the *pushforward* of  $\mu$  under the Markov kernel  $P$  as:

$$P_*(\mu)(B) = \int_S P(s, B) \mu(ds), \quad \forall B \in \mathcal{B}_T.$$

Any measurable function  $f : (S, \mathcal{B}_S) \rightarrow (T, \mathcal{B}_T)$  induces a simple Markov kernel  $s \mapsto \delta_{f(s)}$ . As a result, the usual definition of the pushforward measure (Definition 4.3.5) becomes a special case of the general formulation defined above.

### 4.3.3 Spaces of Measures

The set of all finite measures on a measurable space  $(S, \mathcal{B})$  will be denoted by  $\mathcal{M}(S, \mathcal{B})$ , or simply  $\mathcal{M}S$  when the  $\sigma$ -algebra  $\mathcal{B}$  is clear from context. In particular,  $\mathcal{M}\mathbb{R}$  denotes the Banach space of finite Borel measures on  $\mathbb{R}$ .

$\mathcal{M}S$  forms a real vector space, where addition and scalar multiplication are defined point-wise. Specifically, for  $B \in \mathcal{B}$ ,  $\mu, \nu \in \mathcal{M}S$ , and  $\alpha \in \mathbb{R}$ , the operations are given by

$$(\mu + \nu)(B) = \mu(B) + \nu(B), \quad (\alpha\mu)(B) = \alpha\mu(B).$$

$\mathcal{M}S$  is also a normed space equipped with the total variation norm.

**Definition 4.3.11.** For a measure  $\mu$ , the total variation norm is defined as

$$\|\mu\|_{TV} := \sup \left\{ \sum_{i=1}^n |\mu(B_i)| : \{B_1, \dots, B_n\} \text{ is a finite measurable partition of } S \right\}.$$

For positive measures, this reduces to  $\mu(S)$ , and for probability measures, the norm is 1. The total variation norm turns  $\mathcal{M}_S$  into a Banach space, meaning it is complete under this norm. The following alternative definition is useful to compute the total variation norm between measures.

**Definition 4.3.12.** [47] Let  $P$  and  $Q$  be two probability measures on  $\mathcal{M}_{\mathbb{R}}$ . Define

$$\nu = P + Q, \quad p = \frac{dP}{d\nu}, \quad q = \frac{dQ}{d\nu},$$

where  $\frac{dP}{d\nu}$  denotes the Radon–Nikodym derivative of the measure  $P$  with respect to the measure  $\nu$ . It holds that:

$$\|P - Q\| = \sup_{A \in \mathcal{B}} \left\{ \left| \int_A (p - q) d\nu \right| \right\}.$$

## 4.4 Examples





## Chapter 5

# Quantum computation

The first two sections of this chapter present the mathematical and quantum computing preliminaries necessary for understanding the theory of quantum computation. The introduction to quantum computing is primarily based on [12, 48], while the mathematical foundations are also based on [49] and [50].

The following section introduces core concepts from functional analysis, which are essential in  $C^*$ - and  $W^*$ -algebras. This part is based on [49–52].

Faltam as  $w^*$  algebras

## 5.1 Linear Algebra

It is impossible to present the theory of quantum computation without introducing some concepts of linear algebra within finite-dimensional spaces. This section provides a brief overview of the aspects of linear algebra that are most pertinent to the study of quantum computation.

### 5.1.1 Vector spaces

The basic objects of linear algebra are vector spaces. The vector spaces of interest in this work are the real and complex vector spaces, such as  $\mathbb{R}$  and  $\mathbb{C}^n$  and  $\mathbb{C}^{n \times m}$ , respectively.

**Definition 5.1.1.** A *vector space* (over a field  $\mathcal{F}$ ) consists of a set  $V$  whose elements are called vectors, together with two operations:

- An operation called vector addition that takes two vectors  $v, w \in V$ , and results in a third vector, written  $v + w \in V$ ;

- An operation called scalar multiplication that takes a scalar  $a \in \mathcal{F}$  and a vector  $v \in V$ , and results in a new vector, written  $a \cdot v \in V$

which satisfy the following axioms, for all  $u, v, w \in V$  and  $a_1, a_2 \in \mathcal{F}$ :

1. Vector addition is commutative:  $u + v = v + u$ ;
2. Vector addition is associative:  $(u + v) + w = u + (v + w)$ ;
3. There is a zero vector  $0 \in V$  such that  $v + 0 = v$  for all  $v \in V$ ;
4. Each  $v \in V$  has an additive inverse  $w \in V$  such that  $w + v = 0$ ;
5. Scalar multiplication distributes over scalar addition,  $(a_1 + a_2) \cdot v = a_1 \cdot v + a_2 \cdot v$ ;
6. Scalar multiplication distributes over vector addition,  $a_1 \cdot (v + w) = a_1 \cdot v + a_1 \cdot w$ ;
7. Ordinary multiplication of scalars associates with scalar multiplication,  $(a_1 a_2) \cdot v = a_1 \cdot (a_2 \cdot v)$ ;
8. Multiplication by the scalar 1 is the identity operation,  $1 \cdot v = v$ .

The letters  $\mathcal{R}, V, W$  will often be used to refer to vector spaces. In this section, all vector spaces are assumed to be finite dimensional, unless otherwise stated.

**Definition 5.1.2.** A set of non-zero vectors  $v_1, \dots, v_n$  are *linearly dependent* if there exists a set of complex numbers  $a_1, \dots, a_n$  with  $a_i \neq 0$  for at least one value of  $i$ , such that

$$a_1 v_1 + a_2 v_2 + \dots + a_n v_n = 0.$$

**Definition 5.1.3.** A set of vectors  $v_1, \dots, v_n$  are *linearly independent* if they are not linearly dependent.

**Definition 5.1.4.** A *basis* for a vector space is a sequence of vectors that is linearly independent and that spans the space.

**Definition 5.1.5.** The number of elements in the basis is defined to be the *dimension* of  $V$ , denoted  $\dim(V)$ .

### 5.1.2 Linear operators

**Definition 5.1.6.** A linear operator between vector spaces  $V$  and  $W$  is defined to be any function  $A : V \rightarrow W$  which is linear in its inputs, i.e.

$$A \left( \sum_i a_i v_i \right) = \sum_i a_i A(v_i)$$

Usually  $A(v)$  is just denoted  $Av$ .

Suppose  $V, W$ , and  $R$  are vector spaces, and  $A : V \rightarrow R$  and  $B : W \rightarrow R$  are linear operators. Then the notation  $BA$  is used to denote the *composition* of  $B$  with  $A$ , defined by  $(BA)(v) \equiv B(A(v))$ . Once again,  $(BA)(v)$  is abbreviated as  $BAv$ .

For any choice of complex spaces  $V \in \mathbb{C}^n$  and  $W \in \mathbb{C}^m$ , there is a bijective linear correspondence between the set of operators from  $V$  to  $W$  and the set of  $n \times m$  matrices. The claim that the matrix  $A \in \mathbb{C}^{m \times n}$  is a linear operator just means that

$$A \left( \sum_i a_i v_i \right) = \sum_i a_i A(v_i)$$

is true as an equation where the operation is matrix multiplication of  $A$  by a column vector in  $\mathbb{C}^n$ . Clearly, this is true! On the other hand, suppose  $A : V \rightarrow W$  is a linear operator between vector spaces  $V$  and  $W$ , such that  $V \in \mathbb{C}^n$  and  $W \in \mathbb{C}^m$ . Suppose  $v_1, \dots, v_n$  is a basis for  $V$  and  $w_1, \dots, w_m$  is a basis for  $W$ . Then for each  $j$  in the range  $1, \dots, m$ , there exist complex numbers  $A_{1j}$  through  $A_{nj}$  such that

$$Av_j = \sum_i A_{ij} w_i.$$

The matrix whose entries are the values  $A_{ij}$  is said to form a *matrix representation* of the operator  $A$ . This matrix representation of  $A$  is completely equivalent to the operator  $A$ . As a result, when considering operators on vector spaces of the form  $\mathbb{C}^n$  it is common to refer to the operator  $A$  and its matrix representation interchangeably.

### 5.1.3 Inner product

A linearidade é no segundo ou no primeiro argumentos, é que uns livros usam no primeiro e outros no segundo?

**Definition 5.1.7.** The *inner product*  $\langle \cdot, \cdot \rangle$  is a function from a vector space  $V$  to the field of scalars,  $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathcal{F}$ , that satisfies the following properties for all  $v, w, w_1, \dots, w_n \in V$  and  $a_1, \dots, a_n \in \mathcal{F}$

1. Linearity in the second argument,

$$\left\langle v, \sum_{i=1}^n a_i w_i \right\rangle = \sum_{i=1}^n a_i \langle v, w_i \rangle.$$

2.  $\langle v, w \rangle = \overline{\langle w, v \rangle}$ , where  $\overline{(-)}$  is the complex conjugate operation.

3.  $\langle v, w \rangle \geq 0$  with equality if and only if  $v = 0$ .

The inner product  $\langle v, w \rangle$  of two vectors  $v = (a_1, \dots, a_n), w = (b_1, \dots, b_n) \in \mathbb{C}^n$  is defined as

$$\langle v, w \rangle = \sum_i \bar{a}_i b_i.$$

## Trace

In order to define inner product of a matrix, it is necessary to first define the trace of a matrix.

**Definition 5.1.8.** The trace of a square matrix  $A \in \mathbb{C}^{n \times n}$  defined to be the sum of its diagonal elements,

$$\text{tr}(A) = \sum_i A_{ii}.$$

Alternatively, the trace can also be expressed as:

$$\text{tr}(A) := \sum_{i=1}^n \langle A v_i, v_i \rangle,$$

where  $\{v_i\}_{i=1}^n$  is any orthonormal basis of  $\mathbb{C}^n$ . where  $\{v_i\}$  is an orthonormal basis for  $\mathbb{C}^n$ .

The trace is *cyclic*, that is,  $\text{tr}(AB) = \text{tr}(BA)$ , and *linear*,  $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$ ,  $\text{tr}(a \cdot A) = a \cdot \text{tr}(A)$ , where matrices  $A, B \in \mathbb{C}^{n \times n}$ , and  $a$  is a complex number.

By means of the trace, one defines the inner product of two operators  $A, B \in \mathbb{C}^{m \times n}$  as follows

$$\langle A, B \rangle = \text{tr}(A^\dagger B).$$

In the *finite* dimensional complex vector spaces relevant to quantum computation and quantum information, a *Hilbert space* is equivalent to an inner product space. As a result, both  $\mathbb{C}^n$  and  $\mathbb{C}^{n \times m}$  are Hilbert spaces.

### 5.1.4 Norm and normed spaces

**Definition 5.1.9.** A norm  $\| \cdot \|$  is a function that associates an element of a vector space  $V$  with a non-negative real number, such that the following properties hold:

1. Positive definiteness:  $\|v\| \geq 0$  for all  $v \in V$ , with  $\|v\| = 0$  if and only if  $v = 0$ ;
2. Positive scalability:  $\|av\| = |a|\|v\|$  for all  $v \in V$  scalar  $a$ ;
3. The triangle inequality:  $\|v + w\| \leq \|v\| + \|w\|$  for all  $v, w \in V$ .

**Definition 5.1.10.** A vector space together with a norm is called a *normed vector space*.

Every normed space may be regarded as a metric space, in which the distance  $d(x, y)$  between vectors  $x$  and  $y$  is  $\|x - y\|$ . The relevant properties of  $d(x, y)$  are

1.  $0 < d(x, y) < \infty$  for all  $x$  and  $y$ ,
2.  $d(x, y) = 0$  if and only if  $x = y$ ,
3.  $d(x, y) = d(y, x)$  for all  $x$  and  $y$ ,
4.  $d(x, z) < d(x, y) + d(y, z)$  for all  $x, y, z$ .

Every inner product space is a normed space, where the norm of a vector  $v$  is defined as  $\|v\| = \sqrt{\langle v, v \rangle}$ .

**Definition 5.1.11.** Two vector  $u, v$  are said to be *orthogonal* if  $\langle v, u \rangle = 0$ . An *orthogonal set* is a set of orthogonal vectors of the same vector space.

**Definition 5.1.12.** A *unit vector* is a vector  $v$  such that  $\|v\| = 1$ . It is also said that  $v$  is *normalized* if  $\|v\| = 1$ .

**Definition 5.1.13.** An orthogonal set of unit vectors is called an *orthonormal set*, and when such a set forms a basis it is called an *orthonormal basis*.

### 5.1.5 Eigenvectors and eigenvalues

**Definition 5.1.14.** An *n-permutation* is a function on the first  $n$  positive integers  $\pi = \{1, \dots, n\} \rightarrow \{1, \dots, n\}$  that is one-to-one and onto. In a permutation each number  $1, \dots, n$  appears as output for one and only one input. The *sign* of a permutation  $\text{sgn}(\pi)$  is  $-1$  if the number of inversions in  $\pi$  is odd and is  $+1$  if the number of inversions is even.

**Definition 5.1.15.** The *determinant* of a square matrix  $A \in \mathbb{C}^{n \times n}$  is defined as

$$\det(A) = \sum_{\pi \in S_n} \text{sgn}(\pi) \prod_{i=1}^n A_{i\sigma(i)},$$

Here  $S_n$  is the set of all  $n$ -permutations  $\pi = \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ , and  $\text{sgn}(\pi)$  denotes the sign of the permutation  $\pi$ .

**Definition 5.1.16.** An *eigenvector* of a linear operator  $A$  on a vector space is a non-zero vector  $v$  such that  $Av = \lambda v$ , where  $\lambda$  is a complex number known as the *eigenvalue* of  $A$  corresponding to  $v$ .

The *characteristic polynomial* of a square operator  $A$  is the polynomial  $p(\lambda) = \det(A - \lambda \text{id})$ , where  $\text{id}$  is the identity operator  $\text{id}_V : V \rightarrow V, v \mapsto v$ , where  $V$  is a vector space. The subscript will be omitted unless ambiguity arises. It can be shown that the characteristic function depends only upon the operator  $A$ , and not on the specific matrix representation used for  $A$ . By the fundamental theorem of algebra, every polynomial has at least one complex root, so every operator  $A$  has at least one eigenvalue, and a corresponding eigenvector. The solutions of the *characteristic equation*  $c(\lambda) = 0$  are the eigenvalues of the operator  $A$ . A *diagonal representation* of an operator  $A$  on a vector space  $V$  is an expression of the form  $A = \sum_i \lambda_i v_i v_i^\dagger$ , where the vectors  $v_i$  form an orthonormal set of eigenvectors for  $A$ , with corresponding eigenvalues  $\lambda_i$ , and  $(-)^{\dagger}$  is the adjoint operation.

### 5.1.6 Spectral theorem

**Theorem 5.1.17.** [48] Every normal operator  $A \in \mathbb{C}^{n \times n}$  can be expressed as a linear combination  $\sum_i \lambda_i b_i b_i^\dagger$  where the set  $\{b_i, \dots, b_n\}$  is an orthonormal basis on  $\mathbb{C}^n$ .

Using this last result any function  $f : \mathbb{C} \rightarrow \mathbb{C}$ , can be extended to normal matrices via,

$$f(A) = \sum_i f(\lambda_i) b_i b_i^\dagger \tag{5.1}$$

where  $A = \sum_i \lambda_i b_i b_i^\dagger$  is the spectral decomposition of  $A$ .

### 5.1.7 Important classes of operators/matrices

Linear operators mapping a complex space  $\mathbb{C}^n$  or  $\mathbb{C}^{n \times n}$  to itself will be called *square operators* due to the fact that their matrix representations are square matrices. Therefore, those definitions given in the context of square operators are also valid for square matrices.

The following classes of operators are of particular interest in quantum information theory.

**Definition 5.1.18.** *Normal operators.* A square operator  $A$  is *normal* if  $AA^\dagger = A^\dagger A$ .

**Definition 5.1.19.** *Hermitian operators.* A square operator  $A$  is *hermitian* if  $A = A^\dagger$ . Every Hermitian operator is a normal operator.

**Definition 5.1.20.** *Positive (semidefinite) operators.* A square operator  $A : \mathbb{C}^n \rightarrow \mathbb{C}^n$  is *positive*, denoted  $A \geq 0$ , if  $\langle v, Av \rangle \geq 0$  for all  $v \in \mathbb{C}^n$ . The sum of two positive semidefinite operators is positive semidefinite. Positive matrices are hermitian, and consequently, by the spectral decomposition, have diagonal representation  $A = \sum_i \lambda_i v_i v_i^\dagger$ , with non-negative eigenvalues  $\lambda_i$  [48].

**Definition 5.1.21.** *Unitary operators.* A square operator  $U$  is *unitary* if  $U^\dagger U = UU^\dagger = \text{id}$ . The letter  $U$  will often be used to refer to unitary operators.

Geometrically, unitary operators are important because they preserve inner products between vectors,  $\langle Uv, Uw \rangle = \langle v, w \rangle$  for any two vectors  $v$  and  $w$ . Let  $S_1 = \{v_i\}$  be any orthonormal basis set. Define  $S_2 = \{w_i\} = \{Uv_i | v_i \in S_1\}$ , so  $S_2$  is also an orthonormal basis set, since unitary operators preserve inner products. Note that  $U = \sum_i w_i v_i^\dagger$ . Conversely, if  $\{v_i\}$  are any two orthonormal bases, then it is easily checked that the operator  $U$  defined by  $U = \sum_i w_i^\dagger v_i$  is unitary.

**Definition 5.1.22.** *Density operator.* Positive semidefinite operators that have trace equal to 1 are called *density operators*. By convention, such operators are denoted by the lowercase Greek letter  $\rho$ , often accompanied by subscripts.

**Definition 5.1.23.** *Isometries.* An operator  $A : \mathbb{C}^n \rightarrow \mathbb{C}^m$  is an *isometry* if  $\|Av\| = \|v\|$  for all elements all elements  $v \in \mathbb{C}^n$ .

**Definition 5.1.24.** *Projectors.* A positive semidefinite operator  $P$  is a *projector* if  $P^2 = P$ . Equivalently, a projection operator is a Hermitian operator whose only eigenvalues are 0 and 1. The *orthogonal complement* of  $P$  is the operator  $Q = \text{id} - P$ .

**Definition 5.1.25.** *Diagonal operators.* A square operator  $A$  is *diagonal* if  $A_{ij} = 0$  for all  $i \neq j$ .

### 5.1.8 Useful norms

**Definition 5.1.26.** The *euclidean norm*,  $\|\cdot\|_2$ , of a vector  $v \in \mathbb{C}^n$  is defined as:

$$\|v\|_2 = \sqrt{\langle v, v \rangle}.$$

**Definition 5.1.27.** The *trace norm*,  $\|\cdot\|_1$ , of a matrix  $A$  is defined as:

$$\|A\|_1 = \text{tr} \sqrt{A^\dagger A}$$

This norm is also known as the Schatten 1-norm. The trace norm induces a metric on the set of density matrices which is defined by  $d(\rho, \sigma) = \|\rho - \sigma\|_1$ .

### 5.1.9 Tensor Products and Direct Sums of Complex spaces

**Definition 5.1.28.** The *direct sum* of two vector spaces  $V$  and  $W$ , denoted  $V \oplus W$ , is the space of all pairs  $(v, w)$  where  $v \in V$  and  $w \in W$ .

The inner product in  $V \oplus W$  is defined as follows:

$$\langle (v_1, w_1), (v_2, w_2) \rangle = \langle v_1, v_2 \rangle + \langle w_1, w_2 \rangle.$$

In this context, we introduce two commonly used norms, defined below.

**Definition 5.1.29.** Let  $x = (x_1, x_2, \dots, x_n) \in \mathbb{C}^n$ . The  $L^1$  norm,  $\|\cdot\|_{L^1}$ , is defined by

$$\|x\|_{L^1} = \sum_{i=1}^n |x_i|.$$

**Definition 5.1.30.** Let  $x = (x_1, x_2, \dots, x_n) \in \mathbb{C}^n$ . The  $L^\infty$  norm,  $\|\cdot\|_{L^\infty}$ , is defined by

$$\|x\|_{L^\infty} = \max_{1 \leq i \leq n} |x_i|.$$

**Definition 5.1.31.** Consider two finite complex spaces  $V$  and  $W$  with respective bases  $(e_1, \dots, e_n)$  and  $(f_1, \dots, f_k)$ . The tensor product of  $V$  and  $W$ , denoted  $V \otimes W$ , is defined as the space generated by the basis of syntactic symbols:

$$(e_1 \otimes f_1, \dots, e_1 \otimes f_k, \dots, e_n \otimes f_1, \dots, e_n \otimes f_k).$$

The tensor product of two elements  $v = \sum_i \lambda_i \cdot e_i$  and  $w = \sum_j \mu_j \cdot f_j$  is:

$$v \otimes w = \sum_{i,j} \lambda_i \mu_j \cdot e_i \otimes f_j.$$



The inner product in  $V \otimes W$  is defined as follows

$$\langle e_i \otimes f_j, e_k \otimes f_l \rangle = \langle e_i, e_k \rangle \langle f_j, f_l \rangle.$$

The tensor product of two vectors  $v = (v_1, \dots, v_n) \in \mathbb{C}^n$  and  $w = (w_1, \dots, w_m) \in \mathbb{C}^m$  is defined as the vector  $v \otimes w \in \mathbb{C}^{nm}$  such that

$$v \otimes w = \begin{pmatrix} v_1 w_1 \\ \vdots \\ v_1 w_m \\ \vdots \\ v_n w_1 \\ \vdots \\ v_n w_m \end{pmatrix}.$$

The tensor product of two operators  $A \in \mathbb{C}^{n \times m}$  and  $B \in \mathbb{C}^{p \times q}$  is defined as the operator  $A \otimes B \in \mathbb{C}^{np \times mq}$  such that

$$(A \otimes B)(v \otimes w) = Av \otimes Bw.$$

The tensor product of two operators  $P : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{m \times m}$  and  $Q : \mathbb{C}^{o \times o} \rightarrow \mathbb{C}^{p \times p}$  is defined as the operator  $P \otimes Q \in \mathbb{C}^{n \times n} \otimes \mathbb{C}^{o \times o} \rightarrow \mathbb{C}^{m \times m} \otimes \mathbb{C}^{p \times p}$  such that

$$(P \otimes Q)(A \otimes B) = P(A) \otimes Q(B).$$

The tensor product of two vector spaces corresponding to the direct sums of other vector spaces  $V = V_1 \oplus \dots \oplus V_n$  and  $W = W_1 \oplus \dots \oplus W_n$  is defined as

$$V \otimes W = V_1 \otimes W_1 \oplus \dots \oplus V_1 \otimes W_n \oplus \dots \oplus V_n \otimes W_1 \oplus \dots \oplus \dots \oplus W_n \otimes V_n.$$

The notation  $(-)^{\otimes n}$  will be used to denote the tensor product of a vector space, vector, or operator with itself  $n$  times.

Considering vector spaces  $V, W, R$ , the following equalities hold:

$$v \otimes (w + r) = v \otimes w + v \otimes r, \text{ for all } v, w, r \in V, W, R;$$

$$(v + w) \otimes r = v \otimes r + w \otimes u, \text{ for all } v, w, r \in V, W, R;$$

$$(av) \otimes (bw) = ab(v \otimes w), \text{ for all } v, w \in V, W \text{ and scalars } a, b.$$

## 5.2 Quantum Computing Preliminaries

The basic unit of information in quantum computation is a *quantum bit* or *qubit* [53]. Just as classical bit, which can be in one of two states, 0 or 1, a qubit also has a state. Qubits are represented using *Dirac notation*, where the ket symbol  $|\psi\rangle$  is used to denote a quantum state  $\psi$ . The corresponding bra symbol  $\langle\psi|$  is used to denote the conjugate transpose of the state  $\psi$ . In this setting, the inner product of two states  $|\psi\rangle$  and  $|\phi\rangle$  is denoted  $\langle\psi|\phi\rangle$  and is the same as  $\langle\psi| |\phi\rangle$ .

**Definition 5.2.1.** Each isolated physical system is associated with a Hilbert space, known as the system's *state space*. The system's state is fully characterized by a *state vector*, which is a unit vector within this state space.

### 5.2.1 The 2-Dimensional Hilbert Space

**Definition 5.2.2.** The *state* of a single qubit is described by a normalized vector in the 2-dimensional Hilbert space  $\mathbb{C}^2$ . The states

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

correspond to the classical states 0 and 1, respectively. These states, known as the *computational basis* states, form an orthonormal basis for this vector space.

**Definition 5.2.3.** Unlike classical bits, a qubit is not restricted to the basis states  $|0\rangle$  and  $|1\rangle$ . It can exist in a linear combination of these states, known as a *superposition*. A general qubit state can be written as

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle,$$

where  $\alpha, \beta \in \mathbb{C}$  are called *amplitudes*, and must satisfy the normalization condition  $|\alpha|^2 + |\beta|^2 = 1$ . The values  $|\alpha|^2$  and  $|\beta|^2$  represent the probabilities of measuring the qubit in the states  $|0\rangle$  and  $|1\rangle$ , respectively.

**Definition 5.2.4.** Any normalized qubit state  $|\psi\rangle$  can be written (up to a global phase) as

$$|\psi\rangle = e^{i\gamma} \left( \cos\left(\frac{\theta}{2}\right) |0\rangle + e^{i\phi} \sin\left(\frac{\theta}{2}\right) |1\rangle \right),$$

where  $\theta, \phi, \gamma \in \mathbb{R}$ . The global phase factor  $e^{i\gamma}$  has no observable effect on the outcome of measurements and is often disregarded. Thus, the state is usually represented as:

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right) |0\rangle + e^{i\phi} \sin\left(\frac{\theta}{2}\right) |1\rangle. \quad (5.2)$$

**Definition 5.2.5.** The above parametrization defines a point on the unit sphere in  $\mathbb{R}^3$ , known as the *Bloch sphere*. The angles  $\theta$  and  $\phi$  represent the polar and azimuthal angles, respectively. Each pure qubit state corresponds to a point on this sphere, with associated *Bloch vector* given by  $(\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)$ .

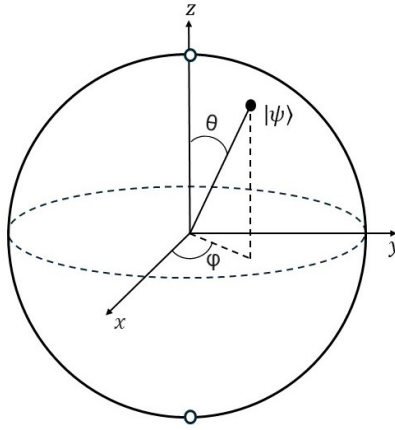


Figure 6: Bloch sphere representation of a qubit

The distance between two quantum states  $|\psi\rangle$  and  $|\psi'\rangle$  is their Euclidean distance in the Bloch sphere [8, 48].

There are infinite points in the Bloch sphere, which might suggest the possibility of encoding an infinite amount of information in the infinite binary expansion of the angle  $\theta$ . However, when a qubit is measured, it collapses to one of the basis states, so only one bit of information can be extracted from a qubit. To accurately determine the amplitudes  $\alpha$  and  $\beta$ , an infinite number of identical qubit copies would need to be measured. Nevertheless, it is still conceptually valid to think of these amplitudes as “hidden information”. One could say that quantum computation is the art of manipulating this hidden information using phenomena such as interference and superposition to perform tasks that would be impossible or inefficient with classical computers.

### 5.2.2 Multi-qubit States

**Definition 5.2.6.** The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. As a result, an  $n$ -qubit state can be represented by a unit vector in  $2^n$ -dimensional Hilbert space,  $\mathbb{C}^{2^n}$ . The notations  $|\psi\rangle \otimes |\phi\rangle$ ,  $|\psi\rangle |\phi\rangle$ , and  $|\psi\phi\rangle$  are used to denote the tensor product of two states  $|\psi\rangle$  and  $|\phi\rangle$ . As for any complex vector,  $|\psi\rangle^{\otimes n}$  denotes the  $n$ -fold tensor product of state  $|\psi\rangle$  with itself. The computational basis states of an  $n$ -qubit system are of the form  $|x_1 \dots x_n\rangle$  and so a quantum state of such a system is specified by  $2^n$  amplitudes. For instance, a two-qubit state can be written as

$$|\psi\rangle = \alpha_{00} |00\rangle + \alpha_{01} |01\rangle + \alpha_{10} |10\rangle + \alpha_{11} |11\rangle.$$

It should be noted that unfortunately, no simple generalization of the Bloch sphere known for multiple qubits.

### Entanglement

**Definition 5.2.7.** An interesting aspect of multi-qubit states is the phenomenon of *entanglement*. This term means strong intrinsic correlations between two (or more) particles when the quantum state of each of them cannot be described independently of the state of the other, i.e. cannot be written as a product of states of the individual qubits. Measuring one qubit of the entangled pair affects the state of the other qubit. This must happen even if the particles are far apart.

In order to better understand this concept, consider the follow *Bell state* or *EPR pair*:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

Upon measuring the first qubit, there are two possible outcomes: 0 with probability  $1/2$  and 1 with probability  $1/2$ . If the first qubit is measured to be 0, the second qubit will also be 0 with probability 1. If the first qubit is measured to be 1, the second qubit will also be 1 with probability 1. Therefore, the measurement outcomes are correlated.

These correlations prompted Einstein, Podolsky, and Rosen to publish a paper [54] questioning the completeness of quantum mechanics in 1935. The EPR paradox presented a dilemma: the existence of entanglement (i.e., correlations that persist regardless of distance) versus local realism and hidden variables. Einstein argued that if two objects, which have interacted in

the past but are now separated, exhibit perfect correlation, they must possess a set of properties determined before their separation. These properties would persist in each object, dictating the outcomes of measurements on both sides. Einstein believed that the strong correlations predicted by quantum mechanics necessitate the existence of additional properties not accounted for by the quantum formalism that determine the measurement results. Therefore, he argued that quantum mechanics might require supplementation, as it may not represent a complete or ultimate description of reality.

In 1964, John Bell made a remarkable discovery: the measurement correlations in the Bell state are stronger than those that could ever occur between classical systems [55]. He explored the idea that each entangled particle might possess hidden properties — unaccounted for by quantum mechanics—that determine the measurement outcomes. Then, through mathematical reasoning, Bell demonstrated that the correlations predicted by any local hidden variable theory cannot exceed a specific level. There is an upper limit of correlations fixed by what today is called the “Bell inequalities”. He found that quantum theory sometimes predicts correlations that exceed this limit. Consequently, an experiment could settle the debate by testing whether or not correlations surpass the bounds he had found following Einstein’s position.

In 1982, Alain Aspect conducted an experiment that confirmed the violation of the Bell inequalities [56]. In this experiment, polarizers were placed more than twelve meters apart. This meant that the correlation obtained could not be explained by the fact that the particles carry within them unmeasured properties. Moreover, it proved that the outcome of the measurement is not determined until the moment of measurement. There seemed to be an instantaneous exchange between two particles at the time of measurement when they were twelve meters apart.

Sixteen years later, Nicolas Gisin [57] and Anton Zeilinger [58] conducted similar experiments, demonstrating that entanglement persists over distances of several kilometers. More recently, [59] extended these tests using entangled photon pairs sent from a satellite to verify Bell’s inequalities over a distance of one thousand kilometers, further confirming that, regardless of the distance, entangled particles behave as an indivisible, inseparable whole. The connection between them is so profound that it appears to challenge the principles of relativity. This phenomenon is known as *quantum nonlocality*.

## 5.2.3 Unitary operators and Measurements

### Pauli Matrices

**Definition 5.2.8.** The Pauli matrices are a set of three  $2 \times 2$  hermitian matrices that are defined as follows:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The eigenvectors and eigenvalues of the Pauli matrices are as follows:

$$\begin{aligned} \sigma_x \begin{pmatrix} 1 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix}, & \sigma_y \begin{pmatrix} 1 \\ i \end{pmatrix} &= \begin{pmatrix} 1 \\ i \end{pmatrix}, & \sigma_z \begin{pmatrix} 1 \\ 0 \end{pmatrix} &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \sigma_x \begin{pmatrix} 1 \\ -1 \end{pmatrix} &= -\begin{pmatrix} 1 \\ -1 \end{pmatrix}, & \sigma_y \begin{pmatrix} 1 \\ -i \end{pmatrix} &= -\begin{pmatrix} 1 \\ -i \end{pmatrix}, & \sigma_z \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= -\begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

The normalized eigenvectors of  $\sigma_x$  are  $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$  and  $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ , and normalized eigenvectors of  $\sigma_y$  are  $|+i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$  and  $|-i\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle)$ . The eigenvectors of  $\sigma_z$  are  $|0\rangle$  and  $|1\rangle$ . These eigenvectors correspond to the  $\hat{x}$ ,  $\hat{y}$  and  $\hat{z}$  axes of the Bloch sphere in [Figure 6](#), respectively.

When matrices  $\sigma_x$ ,  $\sigma_y$  or  $\sigma_z$  are applied to a state on the Bloch sphere, they rotate the state by  $\pi$  radians around the  $\hat{x}$ ,  $\hat{y}$  or  $\hat{z}$  axis, respectively. For example, the action of  $\sigma_x$  on the state  $|0\rangle$  is to rotate it to  $|1\rangle$ , and vice versa. Note that for the eigenstates of these matrices with eigenvalue  $-1$ , this still applies if considering a global phase of  $-1 = e^{i\pi}$ , given that two quantum states  $|\psi\rangle$  and  $e^{i\phi}|\psi\rangle$  are indistinguishable by any quantum measurement.

Matrices  $\sigma_x$  and  $\sigma_z$  will also be referred to as  $X$  and  $Z$ , respectively.

### Unitary operators

**Definition 5.2.9.** *Closed systems*, i.e., systems that do not interact with other systems evolve according to unitary operators. In quantum computation, these unitary operators are also known as *gates*. For a state  $|\psi\rangle$ , a *unitary operator*  $U$  describes an evolution from  $|\psi\rangle$  to  $U|\psi\rangle$ .

**Example 5.2.10.** Pauli matrices are examples of unitary operators. The  $X$  and  $Z$  gates are often referred to as the *not* and *phase flip* gates, respectively. Other important unitary operators include *Hadamard gate*, denoted  $H$ , which maps  $|0\rangle$  to  $|+\rangle$  and  $|1\rangle$  to  $|-\rangle$ , and the

phase-shift gate, denoted  $P$ , which leaves  $|0\rangle$  unaltered applies a phase shift of  $\theta$  to the state  $|1\rangle$ :

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}.$$

When the Pauli matrices are exponentiated, they result in three valuable classes of unitary matrices, corresponding to the rotation operators around the  $\hat{x}$ ,  $\hat{y}$ , and  $\hat{z}$  axes, which are defined as follows:

$$\begin{aligned} R_x(\theta) &= e^{-i\theta\sigma_x/2} = \cos\left(\frac{\theta}{2}\right) \text{id} - i\sin\left(\frac{\theta}{2}\right) \sigma_x = \begin{pmatrix} \cos(\frac{\theta}{2}) & -i\sin(\frac{\theta}{2}) \\ -i\sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{pmatrix}, \\ R_y(\theta) &= e^{-i\theta\sigma_y/2} = \cos\left(\frac{\theta}{2}\right) \text{id} - i\sin\left(\frac{\theta}{2}\right) \sigma_y = \begin{pmatrix} \cos(\frac{\theta}{2}) & -\sin(\frac{\theta}{2}) \\ \sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{pmatrix}, \\ R_z(\theta) &= e^{-i\theta\sigma_z/2} = \cos\left(\frac{\theta}{2}\right) \text{id} - i\sin\left(\frac{\theta}{2}\right) \sigma_z = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix}. \end{aligned}$$

**Theorem 5.2.11.** [48] Suppose  $U$  is a unitary operation on a single qubit. Then there exist real numbers  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  such that

$$U = e^{i\alpha} R_z(\beta) R_y(\gamma) R_z(\delta).$$

**Example 5.2.12.** There are also multi-qubit gates, such as the *controlled-not* gate, denoted  $CNOT$ , which leaves the states  $|00\rangle$  and  $|01\rangle$  unchanged, and maps  $|10\rangle$  and  $|11\rangle$  to each other:

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

In this case, as the state of the first qubit determines if the  $X$  gate is applied to the second qubit, the first qubit is called the *control qubit* and the second qubit the *target qubit*.

There is an “extension” of the controlled-not gate, the controlled- $U$  gate, where  $U$  is a unitary gate acting on a single qubit. This gate applies the gate  $U$  to the target qubit if the control qubit is in state  $|1\rangle$  and does nothing otherwise. It is defined as:

$$\begin{aligned} CU(|0\rangle \otimes |\psi\rangle) &= |0\rangle \otimes |\psi\rangle \\ CU(|1\rangle \otimes |\psi\rangle) &= |1\rangle \otimes U|\psi\rangle. \end{aligned}$$

It should be noted that no completely closed systems exist in the universe. Nevertheless, for many systems, the approximation of a closed system is valid.

## Measurements

There are times when it necessary to observe the system to extract information. This interaction leaves the system no longer closed and, consequently, the evolution of the system is no longer unitary.

**Definition 5.2.13.** The act of measuring a qubit is represented by a set of operators called *measurement operators*, denoted  $\{M_m\}$ . These operators act on the state space of the system being measured. The index  $m$  refers possible measurement outcomes. These measurement operators must satisfy the completeness equation  $\sum_m M_m^\dagger M_m = \text{id}$ , which ensures that the probabilities of all possible outcomes sum to 1. If a measurement  $M_m$  is performed on a state  $|\psi\rangle$  the outcome  $m$  is observed with probability  $p_m = \langle\psi| M_m^\dagger M_m |\psi\rangle$  for each  $m$ . Moreover, after a measurement yielding outcome  $m$ , the state collapses to

$$|\psi'\rangle = \frac{M_m |\psi\rangle}{\sqrt{p_m}}.$$

**Example 5.2.14.** In the case of the computational basis, the measurement operators are the projectors onto the basis states  $|0\rangle$  and  $|1\rangle$  denoted by  $M_0 = |0\rangle\langle 0|$  and  $M_1 = |1\rangle\langle 1|$ , respectively. Considering an arbitrary state  $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ , the probabilities of measuring 0 and 1 are  $p_0 = \langle\psi| M_0 M_0^\dagger |\psi\rangle = \langle\psi| M_0 |\psi\rangle = |\alpha|^2$ , and  $p_1 = \langle\psi| M_1 M_1^\dagger |\psi\rangle = \langle\psi| M_1 |\psi\rangle = |\beta|^2$ , respectively. Consequently the states after measurement are

$$\begin{aligned} |\psi'\rangle &= \frac{M_0 |\psi\rangle}{|\alpha|} = \frac{\alpha}{|\alpha|} |0\rangle = |0\rangle \text{ (with } p = p_0) \quad \text{and} \\ |\psi''\rangle &= \frac{M_1 |\psi\rangle}{|\beta|} = \frac{\beta}{|\beta|} |1\rangle = |1\rangle \text{ (with } p = p_1) \end{aligned}$$

From now on, unless stated otherwise, any reference to measurement should be understood as pertaining to the computational basis.

As previously mentioned, any states  $|\psi\rangle$  and  $e^{i\phi}|\psi\rangle$  are indistinguishable by any quantum measurement. Consider a measurement operator  $M_m$ , the probabilities of obtaining outcome  $m$  are  $\langle\psi| M_m^\dagger M_m |\psi\rangle$  and  $\langle\psi| e^{-i\theta} M_m^\dagger M_m e^{i\theta} |\psi\rangle = \langle\psi| M_m^\dagger M_m |\psi\rangle$ . For this reason, it is said that these states are equal from an observational point of view.

### 5.2.4 Density operators

Until now the state vector formalism was used. However there is an alternative formulation using density operators. The density operator is often known as the *density matrix*, the two terms will be used interchangeably.



**Definition 5.2.15.** A quantum state  $|\psi\rangle$  is said to be a *pure state* if it is completely known, i.e. if it can be written as a ket. In this case, the state can be written in the density operator formalism as  $\rho = |\psi\rangle \langle\psi|$ .

**Definition 5.2.16.** A state that is a probabilistic mixture of pure states is designated a *mixed state*. A mixed state  $\sum_i \alpha_i |\psi_i\rangle$  can be represented by a density operator  $\rho = \sum_i |\alpha_i|^2 |\psi_i\rangle \langle\psi_i|$ . Note that  $|\alpha_i|^2$  is the probability of the system being in state  $|\psi_i\rangle$ .

**Definition 5.2.17** (Unitary Evolution of a Density Operator). When a unitary operator  $U$  is applied to a quantum state described by a density matrix  $\rho$ , the resulting state is given by  $\rho' = U\rho U^\dagger$ .

**Definition 5.2.18** (Measurement of a Density Operator). Given a collection of measurement operators  $\{M_m\}$ , the probability of obtaining outcome  $m$  when measuring a state  $\rho$  is  $p_m = \text{Tr}(M_m \rho M_m^\dagger)$ . After observing outcome  $m$ , the post-measurement state collapses to:

$$\rho' = \frac{M_m \rho M_m^\dagger}{\text{Tr}(M_m \rho M_m^\dagger)}.$$

**Definition 5.2.19.** In [subsection 5.2.1](#) it was shown how to determine the cartesian coordinates of a pure state in the Bloch sphere from the state vector. For an arbitrary  $2 \times 2$  density matrix, the following holds

$$\rho = \frac{1}{2}(\text{id} + r_x \sigma_x + r_y \sigma_y + r_z \sigma_z), \quad (5.3)$$

where  $r = (r_x, r_y, r_z)$  is a real three-dimensional vector such that  $\|r\|_2 \leq 1$ . This vector is known as the *Bloch vector for the state*  $\rho$ . Since  $\rho$  is Hermitian,  $r_x, r_y$  and  $r_z$  are always real. The inverse map of [Equation 5.3](#) is

$$r_\mu = \text{Tr}(\rho \sigma_\mu) \quad (5.4)$$

Note that given that the trace is linear and matrix multiplication distributes over matrix addition, the cartesian coordinates of an operator consisting of the sum or subtraction of density operators can also be determined by [Equation 5.4](#).

## Reduced density operator

Density operators are particularly well-suited for describing individual subsystems of a composite quantum system. This type of description is provided by the *reduced density operator*

**Definition 5.2.20.** Let  $|a_1\rangle, |a_2\rangle$  be vectors in the Hilbert space of system  $A$ , and  $|b_1\rangle, |b_2\rangle$  vectors in the Hilbert space of system  $B$ . The *partial trace over  $B$*  is defined by:

$$\begin{aligned}\text{Tr}_B(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) &= |a_1\rangle\langle a_2| \text{Tr}(|b_1\rangle\langle b_2|) \\ &= |a_1\rangle\langle a_2| \sum_{\mu} \langle\mu| |b_1\rangle\langle b_2| |\mu\rangle \\ &= |a_1\rangle\langle a_2| \sum_{\mu} \langle\mu| b_1\rangle\langle b_2| \mu\rangle \\ &= |a_1\rangle\langle a_2| \sum_{\mu} \langle b_2| \mu\rangle\langle \mu| b_1\rangle \\ &= |a_1\rangle\langle a_2| \langle b_2| b_1\rangle\end{aligned}$$

where  $\{|\mu\rangle\}$  is an orthonormal basis spanning the state space of  $B$ . In certain situations it is more advantageous to consider the reduced density operator for subsystem  $A$  defined as:

$$\rho_A = \text{Tr}_B(\rho_{AB}) = \sum_{\mu} \langle\mu| \rho_{AB} |\mu\rangle,$$

where  $\{|\mu\rangle\}$ , span the state space of  $B$  and act only in the state space of  $B$ .

**Definition 5.2.21.** Given physical systems  $A$  and  $B$  whose composite system is given by the density operator  $\rho_{AB}$ , the *reduced density operator* for subsystem  $A$  is

$$\rho_A = \text{Tr}_B(\rho_{AB}).$$

Note that, by linearity, the partial trace of  $\rho_{AB} = \sum_{ijkl} p_{ijkl} |a_i\rangle\langle a_j| \otimes |b_k\rangle\langle b_l|$  is

$$\begin{aligned}\text{Tr}_B(\rho_{AB}) &= \sum_{ijkl} p_{ijkl} \text{Tr}_B(|a_i\rangle\langle a_j| \otimes |b_k\rangle\langle b_l|) \\ &= \sum_{ijkl} p_{ijkl} |a_i\rangle\langle a_j| \sum_{\mu} \langle\mu| b_k\rangle\langle b_l| \mu\rangle = \sum_{ijkl} p_{ijkl} |a_i\rangle\langle a_j| \langle b_l| b_k\rangle\end{aligned}$$

Similarly, the *reduced density operator for subsystem  $B$*  is  $\rho_B = \text{Tr}_A(\rho_{AB})$ .

Decidir onde colocar a def de trace-nonincreasing op e quantum operation: aqui ou na parte das contribuições qd falar da categoria do selinger

## 5.2.5 Quantum Channels

Thus far, only two types of quantum operations have been discussed: unitary operators, which describe the evolution of a closed quantum system, and measurements, which describe the act of observing a quantum system. Now, a new type of quantum operation that

accounts for the more realistic notion of interaction between a quantum system and an environment will be introduced. Nonetheless, it is necessary to first introduce a few key concepts.

**Definition 5.2.22.** A super-operator  $Q$  is a linear map between the space of operators on a Hilbert space.

**Definition 5.2.23.** A super-operator  $Q$  is called *positive* if it sends positive matrices to positive matrices, i.e.  $A \geq 0 \Rightarrow QA \geq 0$ .

**Definition 5.2.24.** A super-operator  $Q$  is said to be *completely positive* if for all  $k$ ,

$$Q \otimes I_{\mathbb{C}^{k \times k}} : \mathbb{C}^{n \times n} \otimes \mathbb{C}^{k \times k} \rightarrow \mathbb{C}^{m \times m} \otimes \mathbb{C}^{k \times k}$$

is positive.

**Definition 5.2.25.** A super-operator  $Q$  is called *trace-preserving* if  $\text{Tr}(QA) = \text{Tr}(A)$ .

Since density matrices are positive, any physically allowed transformation must be represented by a positive operator. Nonetheless, this is not sufficient on its own: since one can always extend the space  $\mathbb{C}^{n \times n}$  to  $\mathbb{C}^{n \times n} \otimes \mathbb{C}^{m \times m}$  by adjoining a new quantum system, any physically allowed transformation must be completely positive. Finally, since the trace of a density matrix is always 1, any physically allowed transformation must be trace-preserving.

**Definition 5.2.26.** A **Completely Positive Trace-Preserving (CPTP)** operator is traditionally called a *quantum channel*.

### Kraus operator sum representation

Assume that there is a quantum system  $S$  of interest which is a subsystem of a larger system which also includes an environment  $E$ . These systems have a joint unitary evolution described by a unitary operator  $U$  acting on the composite system,  $U(\rho_{SE}) = U\rho_{SE}U^\dagger$ . Given that density matrices are positive operators, by (Definition 5.1.20), the density operator of the environment  $\rho_E$  initially can be written as

$$\rho_E = \sum_i p_i |i\rangle \langle i|$$

where  $|i\rangle$  form an orthonormal basis for the state space of  $E$  and  $p_i$  are positive.

The state of the subsystem  $S$  after the unitary evolution corresponds to the partial trace of the joint state over the environment,

$$\begin{aligned}\rho'_S &= \text{Tr}_E(U \rho_{SE} U^\dagger) \\ &= \sum_{\mu} \langle \mu | U \rho_{SE} U^\dagger | \mu \rangle\end{aligned}$$

where  $\{|\mu\rangle\}$  span the state space of  $E$ .

Considering that initially both systems are completely decoupled, the initial state of the system can be written as  $\rho_{SE} = \rho_S \otimes \rho_E$ . Thus,

$$\begin{aligned}\rho'_S &= \sum_{\mu} \langle \mu | U \rho_S \otimes \sum_i p_i | i \rangle \langle i | U^\dagger | \mu \rangle \\ &= \sum_{\mu i} \sqrt{p_i} \langle \mu | U | i \rangle \rho_S \sqrt{p_i} \langle i | U^\dagger | \mu \rangle \\ &= \sum_{\mu i} K_{\mu i} \rho_S K_{\mu i}^\dagger\end{aligned}$$

where the set of operators  $\{K_{\mu i}\}$  is designated *Kraus operators* and  $K_{\mu i} = \sqrt{p_i} \langle \mu | U | i \rangle$ . Note that  $\{|\mu\rangle\}$  and  $\{|i\rangle\}$ , act only in the state space of  $E$ .

**Definition 5.2.27.** The equation  $\rho'_S = \sum_{\mu i} K_{\mu i} \rho_S K_{\mu i}^\dagger$  is called an **Operator Sum Representation (OSR)**. An **OSR** can be thought of as a quantum channel that maps  $\rho_S$  to  $\sum_{\mu i} K_{\mu i} \rho_S K_{\mu i}^\dagger$ , given this map is **CPTP** [12, 60].

### Non-selective measurements

In the previously presented formalism to represent all the possible outcomes of a measurement, described by a set of operators  $\{M_m\}$ , on a state  $\rho$ , it would be necessary to write that state  $\rho$  collapse to state  $\rho_m = \frac{M_m^\dagger \rho M_m}{\text{Tr}(M_m^\dagger \rho M_m)}$  with probability  $p_m = \text{Tr}(M_m^\dagger \rho M_m)$ , for each possible outcome  $m$ . Although the selective description above is useful conceptually, it is often impractical for calculations. Instead, one uses *non-selective measurements*, in which the possible outcomes are not explicitly stated.

**Definition 5.2.28.** A *non-selective measurement* is a quantum measurement in which the post-measurement state of the system is then given by the weighted sum over all possible outcomes:

$$\rho = \sum_m p(m) \rho_m = \sum_m M_m \rho M_m^\dagger.$$

This last equality corresponds to an Kraus operator sum representation, where the set of Kraus operators is  $\{M_m\}$ .

## Norms on quantum channels

**Definition 5.2.29.** The norm of a super-operator  $Q : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{m \times m}$  is defined:

$$\|Q\|_1 = \max\{\|QA\|_1 \mid \|A\|_1 = 1\},$$

where  $A \in \mathbb{C}^{n \times n}$

Unfortunately, this norm is not stable under tensoring, given that the inequation  $\|Q \otimes I_{\mathbb{C}^{n \times n}}\|_1 \leq \|Q\|_1$  does not hold [12]. As a result, the diamond norm, which is based on the trace norm, is used instead in the context of quantum channels.

**Definition 5.2.30.** Given a super-operator  $Q : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{m \times m}$ , the diamond norm,  $\|\cdot\|_\diamond$ , is defined as:

$$\|Q\|_\diamond = \|Q \otimes \text{id}_{\mathbb{C}^{n \times n}}\|_1$$

For this norm it holds that for all super-operators  $Q : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{m \times m}$  and  $S : \mathbb{C}^{m \times m} \rightarrow \mathbb{C}^{o \times o}$ , if  $Q$  is a quantum channel then  $\|SQ\|_\diamond \leq \|S\|_\diamond$  [12, Proposition 3.44 and Proposition 3.48], and if  $S$  is a quantum channel, then  $\|SQ\|_\diamond \leq \|Q\|_\diamond$ . This is a desirable property, as it guarantees that quantum operations do not increase the distance between states, and as a consequence, composition of programs is valid.

Since the diamond norm is generally difficult to compute, we will rely on the following properties, as given by the theorems below:

**Theorem 5.2.31.** [12, Theorem 3.55] Let  $n \leq m$ , let  $V_0, V_1 : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{m \times m}$  be isometries, and define CPTP operators  $\phi_0, \phi_1 : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{m \times m}$  as

$$\phi_0(\rho) = V_0 \rho V_0^\dagger \quad \text{and} \quad \phi_1(\rho) = V_1 \rho V_1^\dagger$$

for all  $\rho \in \mathbb{C}^{n \times n}$ . There exists a unit vector  $u \in \mathbb{C}^{n \times n}$  such that

$$\|\phi_0(uu^\dagger) - \phi_1(uu^\dagger)\|_1 = \|\phi_0 - \phi_1\|_\diamond.$$

**Theorem 5.2.32.** [12, Theorem 3.56] Let  $\phi : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$  be a quantum channel, let  $\varepsilon \in [0, 2]$ , and suppose that

$$\|\phi(\rho) - \rho\|_1 \leq \varepsilon$$

for every density operator  $\rho \in \mathbb{C}^{n \times n}$ . It holds that

$$\|\phi - \text{id}_\sigma\|_1 \leq \sqrt{2\varepsilon}.$$

### 5.2.6 Quantum circuits

As quantum computation remains in its early stages of development, programming is primarily based on the use of *quantum circuits*.

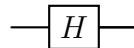
**Definition 5.2.33.** A *quantum circuit* consists of wires and quantum gates, which serve to transmit and manipulate quantum information. Each wire corresponds to a qubit, while the gates represent operations that can be applied to these qubits.

In this subsection the notation for the quantum gates used in this work will be introduced.

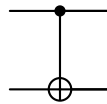
Wires in parallel represent the tensor product of the respective qubits. For instance,  $\psi_0 \otimes \psi_1$  corresponds to

$$\begin{array}{c} |\psi_0\rangle \text{ —————} \\ |\psi_1\rangle \text{ —————} \end{array}$$

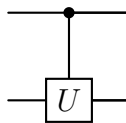
The single bit gates presented in [Section 5.2.3](#) are represented as a box with the symbol of the gate inside. For example, the Hadamard gate is represented as



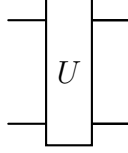
The controlled-not gate, which is a two-qubit gate, is represented as



Similarly, the controlled- $U$  gate, where  $U$  is an unitary single-qubit gate, is represented as

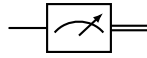


An arbitrary unitary operator acting on  $n$  qubits is represented as a box acting on  $n$  wires. For instance, the operator  $U$  acting on two qubits is represented as



**CPTP** maps are depicted as boxes containing the corresponding map symbols. The relevant **CPTP** operators will be introduced in ??.

The measurement operation is represented by a “meter” symbol. Given that output of a measurement is a classical bit, the wire representing the output of a measurement is a classical wire, represented by a double line.



### 5.2.7 No-cloning theorem

The no-cloning theorem states that it is impossible to duplicate an unknown quantum bit [61]. In this subsection, an elementary proof of this theorem will be presented.

Suppose that there exists a unitary operator  $U$  that receives a qubit  $|\psi\rangle$  and some standard pure state  $|s\rangle$  as input and outputs the state  $|\psi\rangle \otimes |\psi\rangle$ . The action of  $U$  can be written as

$$U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle$$

Consider the application of  $U$  to two pure states  $|\psi_0\rangle$  and  $|\psi_1\rangle$ ,

$$U(|\psi_0\rangle \otimes |s\rangle) = |\psi_0\rangle \otimes |\psi_0\rangle$$

$$U(|\psi_1\rangle \otimes |s\rangle) = |\psi_1\rangle \otimes |\psi_1\rangle.$$

Given that unitary operators preserve inner products, the following equality should hold:

$$\langle\psi_0|\psi_1\rangle = (\langle\psi_0|\psi_1\rangle)^2$$

This equation is only satisfied if  $\langle\psi_0|\psi_1\rangle = 0$  or  $\langle\psi_0|\psi_1\rangle = 1$ . The first case implies that  $|\psi_0\rangle$  and  $|\psi_1\rangle$  are orthogonal, and the second case implies that they are in the same state. Therefore, it is only possible to clone orthogonal states. These are the states perfectly distinguishable by measurement and thus are equivalent to copying classical information. For instance, it is impossible to clone qubits  $\psi_0 = |0\rangle$  and  $\psi_1 = |-\rangle$ , since they are not orthogonal.

It should be noted that this principle is upheld by the type system outlined in Figure 1, which does not allow the repeated use of a variable (seen as a quantum resource).

## 5.3 Functional Analysis

In this section, we are no longer restricted to finite dimensional vector spaces; the term “vector space” now also encompasses infinite-dimensional ones.

Resolver questões de notação, espaços e tensores

### 5.3.1 Hilbert Spaces

Aqui  $\mathcal{H}$  and  $\mathcal{K}$  are typical designations for Hilbert Spaces

**Definition 5.3.1.** A Hilbert space  $\mathcal{H}$  is an inner product space that is complete with respect to the norm induced by the inner product (i.e.  $\|v\| = \sqrt{\langle v, v \rangle}$  for all  $v \in \mathcal{H}$ ).

The letters  $\mathcal{H}, \mathcal{K}$  will often be used to refer to Hilbert spaces.

The definition of trace is extended in this setting as follows:

**Definition 5.3.2.** Let  $\mathcal{H}$  be an Hilbert space and  $0 \leq T \in \mathcal{B}(\mathcal{H})$ . The trace is defined as

$$\text{Tr}(T) := \sum_i \langle T v_i, v_i \rangle \in [0, \infty],$$

where  $\{v_i\}$  is an orthonormal basis for  $\mathcal{H}$ .

**Definition 5.3.3.** Let  $\mathcal{H}$  be an Hilbert space. An operator  $T$  is *trace class* if  $\text{Tr}(|T|) < \infty$ , where  $|T| = (\overline{T}T)^{1/2}$ .

**Definition 5.3.4.** Let  $\mathcal{H}$  and  $\mathcal{K}$  be Hilbert spaces. We denote by  $\mathcal{H} \otimes_2 \mathcal{K}$  the Hilbert space tensor product, obtained by completing  $\mathcal{H} \otimes \mathcal{K}$  equipped with the standard inner product characterized by

$$\langle k_1 \otimes h_1, k_2 \otimes h_2 \rangle = \langle k_1, k_2 \rangle \cdot \langle h_1, h_2 \rangle.$$

### 5.3.2 Topology

Topology is the abstract mathematical study of concepts like convergence and approximation, generalizing familiar notions from calculus and analysis. Note that, for instance, in a metric space, a sequence  $\{x_n\}$  with  $n \in \mathbb{N}$  converges to a point  $x$  if the distance  $d(x_n, x)$  tends to zero; that is, for every  $\varepsilon > 0$ , there exists  $n_0$  such that  $d(x_n, x) < \varepsilon$  for all  $n \geq n_0$ .



However, metric spaces are not sufficient to describe all types of convergence. An example is the pointwise convergence of all real-valued functions on the interval  $[0, 1]$ . In fact, there is no metric on the space of all real functions on the interval  $[0, 1]$  for which one can define a distance function  $d(f_n, f)$  such that  $d(f_n, f) \rightarrow 0$  if and only if  $f_n(x) \rightarrow f(x)$  for every  $x \in [0, 1]$ . A foundational idea in topology is that of a *neighborhood*—a collection of points considered “sufficiently close” to a given point. From this arises the concept of *open sets*, which are sets that serve as neighborhoods for all their points. The collection of all such open sets defines a *topology*, and a set equipped with a topology becomes a *topological space*. This framework introduces some subtleties: for example, traditional sequences are often inadequate for capturing convergence, requiring the more general notion of *nets*, which are indexed over broader structures than the natural numbers.

**Definition 5.3.5.** A topology  $\tau$  on a set  $X$  is a collection of subsets of  $X$  satisfying the following properties:

1.  $\emptyset \in \tau$  and  $X \in \tau$ .
2.  $\tau$  is closed under finite intersections: if  $U_1, U_2, \dots, U_n \in \tau$ , then  $\bigcap_{i=1}^n U_i \in \tau$ .
3.  $\tau$  is closed under arbitrary unions: if  $\{U_\alpha\}_{\alpha \in A} \subseteq \tau$ , then  $\bigcup_{\alpha \in A} U_\alpha \in \tau$ .

A nonempty set  $S$  equipped with a topology  $\tau$  is called a *topological space*, and is denoted by  $(S, \tau)$  (or simply  $X$  when no ambiguity arises). A member of  $\tau$  is called an *open set* in  $S$ . The complement of an open set is a *closed set*.

A set  $S$  can have many different topologies. The family of all topologies on  $S$  is partially ordered by set inclusion. If  $\tau_1 \subset \tau_2$ , that is, if every  $\tau_1$ -open set is also  $\tau_2$ -open, then we say that  $\tau_1$  is *weaker* or *coarser* than  $\tau_2$ , and that  $\tau_2$  is *stronger* or *finer* than  $\tau_1$ .

**Example 5.3.6.** Standard examples of topologies are presented below:

1. *Trivial (or indiscrete) topology:* On a set  $S$ , the trivial topology consists only of the sets  $\emptyset$  and  $X$ . These are also the only closed sets.
2. *Discrete topology:* The discrete topology on a set  $S$  consists of all possible subsets of  $X$ . In this topology, every set is both open and closed.
3. *Standard topology on  $\mathbb{R}$ :* The metric  $d(x, y) = |x - y|$  on  $\mathbb{R}$  induces a topology where open sets are unions of open intervals. This is known as the standard topology on  $\mathbb{R}$ .

**Definition 5.3.7.** The *norm topology* induced by a norm  $\|\cdot\|$  is the metrizable topology generated by the metric  $d(x, y) = \|x - y\|$ .

**Definition 5.3.8.** A *topological vector space* is a vector space  $V$  equipped with a linear  $\tau$  such that:

1. every singleton  $\{v\} \subset V$  is a closed set, and
2. the vector space operations (addition and scalar multiplication) are continuous with respect to  $\tau$ . That is, the addition map  $(x, y) \mapsto x + y$ , from the Cartesian product  $V \times V$  into  $V$ , is continuous, and the scalar multiplication map  $(r, x) \mapsto rx$ , from  $\mathcal{F} \times V$  into  $V$ , is also continuous.

**Definition 5.3.9.** Let  $V$  be a vector space. Linear maps from  $V$  to its scalar field are called *linear functionals*. The set of all continuous linear functionals on  $V$  forms a vector space, called the *(topological) dual space* of  $V$ , and is denoted by  $V^*$ . It is common to designate elements of the dual space  $V^*$  by  $v^*$ .

**Theorem 5.3.10.** Let  $V$  be a normed vector space. For each  $v^* \in V^*$ , define its norm by

$$\|v^*\| := \sup \{|v^*(v)| : \|v\| = 1\}.$$

This defines a norm on  $V^*$  under which  $V^*$  is a Banach space. Moreover, for every  $v \in V$ , we have

$$\|v\| = \sup \{|v^*(v)| : \|v^*\| = 1\}.$$

As a consequence, the map  $v^* \mapsto v^*(v)$  defines a bounded linear functional on  $V^*$ , and its norm equals  $\|v\|$ .

**Definition 5.3.11.** Let  $V$  be a vector space. The *weak\*-topology* on the dual space  $V^*$  is the coarsest topology that makes all evaluation maps

$$v^* \mapsto v^*(v)$$

continuous for every  $v \in V$ .

## 5.4 $W^*$ -Algebras

## 5.5 Categories for (first-order) quantum computation

**Example 5.5.1.** The category CPTP is the category whose objects are natural numbers  $n \geq 1$  and whose morphisms  $n \rightarrow m$  are quantum channels  $C^{n \times n} \rightarrow C^{m \times m}$ .

**Example 5.5.2.** The category CPS is the category whose objects are natural numbers  $n \geq 1$  and whose morphisms  $n \rightarrow m$  are completely positive trace-nonincreasing maps  $C^{n \times n} \rightarrow C^{m \times m}$ .

## 5.6 Examples

Interpretação Teleporte Quântico

Daqui tirar o measurement error



## **Chapter 6**

# **Conclusions and future work**

Conclusions and future work.

### **6.1 Conclusions**

### **6.2 Prospect for future work**



# Bibliography

- [1] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6-7), 1982.
- [2] Peter W Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.
- [3] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018.
- [4] A Robert Calderbank and Peter W Shor. Good quantum error-correcting codes exist. *Physical Review A*, 54(2):1098, 1996.
- [5] Daniel Gottesman. *Stabilizer codes and quantum error correction*. California Institute of Technology, 1997.
- [6] Andrew M Steane. Error correcting codes in quantum theory. *Physical Review Letters*, 77(5):793, 1996.
- [7] Austin G Fowler, Matteo Mariantoni, John M Martinis, and Andrew N Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86(3):032324, 2012.
- [8] Joel J Wallman and Joseph Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Physical Review A*, 94(5):052325, 2016.
- [9] Prakash Murali, Jonathan M Baker, Ali Javadi-Abhari, Frederic T Chong, and Margaret Martonosi. Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers. In *Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems*, pages 1015–1029, 2019.
- [10] Aram W Harrow, Benjamin Recht, and Isaac L Chuang. Efficient discrete approximations of quantum gates. *Journal of Mathematical Physics*, 43(9):4445–4451, 2002.

- [11] Lukas Burgholzer and Robert Wille. Advanced equivalence checking for quantum circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 40(9):1810–1824, 2020.
- [12] John Watrous. *The theory of quantum information*. Cambridge university press, 2018.
- [13] Glynn Winskel. *The formal semantics of programming languages: an introduction*. MIT press, 1993.
- [14] Jianjun Zhao. Quantum software engineering: Landscapes and horizons. *arXiv preprint arXiv:2007.07047*, 2020.
- [15] Manuel A Serrano, Jose A Cruz-Lemus, Ricardo Perez-Castillo, and Mario Piattini. Quantum software components and platforms: Overview and quality assessment. *ACM Computing Surveys*, 55(8):1–31, 2022.
- [16] Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023.
- [17] Benjamin Bichsel, Maximilian Baader, Timon Gehr, and Martin Vechev. Silq: A high-level quantum language with safe uncomputation and intuitive semantics. In *Proceedings of the 41st ACM SIGPLAN Conference on Programming Language Design and Implementation*, pages 286–300, 2020.
- [18] Alexander S Green, Peter LeFanu Lumsdaine, Neil J Ross, Peter Selinger, and Benoît Valiron. Quipper: a scalable quantum programming language. In *Proceedings of the 34th ACM SIGPLAN conference on Programming language design and implementation*, pages 333–342, 2013.
- [19] Krysta Svore, Alan Geller, Matthias Troyer, John Azariah, Christopher Granade, Bettina Heim, Vadym Kliuchnikov, Mariia Mykhailova, Andres Paz, and Martin Roetteler. Q# enabling scalable quantum computing and development with a high-level dsl. In *Proceedings of the real world domain specific languages workshop 2018*, pages 1–10, 2018.
- [20] Margherita Zorzi. On quantum lambda calculi: a foundational perspective. *Mathematical Structures in Computer Science*, 26(7):1107–1195, 2016.
- [21] Frederic T Chong, Diana Franklin, and Margaret Martonosi. Programming languages and compiler design for realistic quantum hardware. *Nature*, 549(7671):180–187, 2017.



- [22] Shih-Han Hung, Kesha Hietala, Shaopeng Zhu, Mingsheng Ying, Michael Hicks, and Xiaodi Wu. Quantitative robustness analysis of quantum programs. *Proceedings of the ACM on Programming Languages*, 3(POPL):1–29, 2019.
- [23] Runzhou Tao, Yunong Shi, Jianan Yao, John Hui, Frederic T Chong, and Ronghui Gu. Gleipnir: toward practical error analysis for quantum programs. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, pages 48–64, 2021.
- [24] Fredrik Dahlqvist and Renato Neves. The syntactic side of autonomous categories enriched over generalised metric spaces. *Logical Methods in Computer Science*, Volume 19, Issue 4, 2023.
- [25] Jean-Yves Girard, Yves Lafont, and Laurent Regnier. *Advances in linear logic*, volume 222. Cambridge University Press, 1995.
- [26] P Nick Benton. A mixed linear and non-linear logic: Proofs, terms and models. In *International Workshop on Computer Science Logic*, pages 121–135. Springer, 1994.
- [27] Hendrik P Barendregt et al. *The lambda calculus*, volume 3. North-Holland Amsterdam, 1984.
- [28] Paul Bernays. Alonzo church. an unsolvable problem of elementary number theory. american journal of mathematics, vol. 58 (1936), pp. 345–363. *The Journal of Symbolic Logic*, 1(2):73–74, 1936.
- [29] Ian Mackie, Leopoldo Román, and Samson Abramsky. An internal language for autonomous categories. *Applied Categorical Structures*, 1(3):311–343, 1993.
- [30] Roy L. Crole. *Categories for Types*. Cambridge University Press, 1994.
- [31] Peter Selinger. Lecture notes on the lambda calculus, 2013.
- [32] Noson S. Yanofsky. *Monoidal Category Theory: Unifying Concepts in Mathematics, Physics, and Computing*. MIT Press, 2024.
- [33] Michael Barr and Charles Wells. *Category Theory for Computing Science*. Prentice Hall, 1990.

- [34] Saunders Mac Lane. *Categories for the working mathematician*, volume 5. Springer Science & Business Media, 2013.
- [35] John W Backus, Friedrich L Bauer, Julien Green, Charles Katz, John McCarthy, Alan J Perlis, Heinz Rutishauser, Klaus Samelson, Bernard Vauquois, Joseph Henry Wegstein, et al. Report on the algorithmic language algol 60. *Communications of the ACM*, 3(5):299–311, 1960.
- [36] Michael Shulman. A practical type theory for symmetric monoidal categories. *Theory and Applications of Categories*, 37(5):863–907, 2021.
- [37] Fredrik Dahlqvist and Renato Neves. A complete v-equational system for graded lambda-calculus. *Electronic Notes in Theoretical Informatics and Computer Science*, 3, 2023.
- [38] Radu Mardare, Prakash Panangaden, and Gordon Plotkin. Quantitative algebraic reasoning. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 700–709, 2016.
- [39] Radu Mardare, Prakash Panangaden, and Gordon Plotkin. On the axiomatizability of quantitative algebras. In *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12. IEEE, 2017.
- [40] Samuel Eilenberg and Saunders MacLane. General Theory of Natural Equivalences. *Transactions of the American Mathematical Society*, 58(2):231–294, 1945.
- [41] G. Gierz, K. H. Hofmann, K. Keimel, J. D. Lawson, M. Mislove, and D. S. Scott. *Continuous Lattices and Domains*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2003.
- [42] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 2 edition, 2002.
- [43] Gilles Barthe, Joost-Pieter Katoen, and Alexandra Silva, editors. *Foundations of Probabilistic Programming*. Cambridge University Press, 2020.
- [44] Fredrik Dahlqvist, Alexandra Silva, and Dexter Kozen. Semantics of Probabilistic Programming: A Gentle Introduction. In Gilles Barthe, Joost-Pieter Katoen, and Alexandra

- Silva, editors, *Foundations of Probabilistic Programming*, pages 1–42. Cambridge University Press, 2020.
- [45] Krishna B. Athreya and Soumendra N. Lahiri. *Measure Theory and Probability Theory*. Springer Science & Business Media, 2006.
  - [46] Charalambos D. Aliprantis and Kim C. Border. Infinite Dimensional Analysis: A Hitchhiker’s Guide. pages 301–330. Springer, 1999.
  - [47] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated, 1 edition, 2008.
  - [48] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.
  - [49] Walter Rudin. *Functional Analysis*. McGraw-Hill, 1991.
  - [50] A Hitchhiker’s Guide. *Infinite dimensional analysis*. Springer, 2006.
  - [51] John B. Conway. *A Course in Functional Analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer, 2007.
  - [52] John B. Conway. *A Course in Operator Theory*. American Mathematical Society, 2000.
  - [53] Simon Perdrix. Quantum entanglement analysis based on abstract interpretation. In *International Static Analysis Symposium*, pages 270–282. Springer, 2008.
  - [54] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.
  - [55] John S Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.
  - [56] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental test of bell’s inequalities using time-varying analyzers. *Physical review letters*, 49(25):1804, 1982.
  - [57] Wolfgang Tittel, Jürgen Brendel, Bernard Gisin, Thomas Herzog, Hugo Zbinden, and Nicolas Gisin. Experimental demonstration of quantum correlations over more than 10 km. *Physical Review A*, 57(5):3229, 1998.

- [58] Jian-Wei Pan, Dik Bouwmeester, Harald Weinfurter, and Anton Zeilinger. Experimental entanglement swapping: entangling photons that never interacted. *Physical review letters*, 80(18):3891, 1998.
- [59] Juan Yin, Yuan Cao, Yu-Huai Li, Sheng-Kai Liao, Liang Zhang, Ji-Gang Ren, Wen-Qi Cai, Wei-Yue Liu, Bo Li, Hui Dai, et al. Satellite-based entanglement distribution over 1200 kilometers. *Science*, 356(6343):1140–1144, 2017.
- [60] Daniel A Lidar. Lecture notes on the theory of open quantum systems. *arXiv preprint arXiv:1902.00967*, 2019.
- [61] William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982.



Place here information about funding, FCT project, etc. in which the work is framed. Leave empty otherwise.