

UNIVERSITY OF CALGARY

An investigation of quantum and reversible computing

by

Brett Gordon Giles

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA

August, 2013

© Brett Gordon Giles 2013

Abstract

Acknowledgements

Table of Contents

Abstract	i
Acknowledgements	ii
Table of Contents	iii
List of Tables	vi
List of Figures	vii
List of Symbols	viii
1 Introduction	4
2 Abstract Computability	5
2.1 Categories	5
2.1.1 Enrichment of categories	7
2.1.2 Examples of categories	7
2.1.3 Properties of maps	9
2.1.4 Limits and colimits in categories	11
2.1.5 Functors and natural transformations	12
2.2 Restriction categories	14
2.2.1 Enrichment and meets	15
2.2.2 Range categories	18
2.2.3 Partial monics, sections and isomorphisms	20
2.2.4 Split restriction categories	22
2.2.5 Partial Map Categories	26
2.2.6 Restriction products and Cartesian restriction categories	27
2.2.7 Graphic Categories	29
3 Inverse categories	33
3.1 Inverse products	33
3.1.1 Inverse categories with restriction products	33
3.1.2 Inverse products	35
3.1.3 Discrete inverse categories	37
3.1.4 The inverse subcategory of a discrete restriction category	43
3.2 Completing a discrete inverse category	46
3.2.1 The restriction category $\widetilde{\mathbb{X}}$	46
3.2.2 The category $\widetilde{\mathbb{X}}$ is a discrete restriction category	55
3.2.3 Equivalence of categories	59
3.2.4 Examples of the $(-)$ construction	64
3.3 Reversible computation	65
3.3.1 Reversible Turing machines	66
3.3.2 Reversible automata and linear combinatory algebras	71
3.4 Coproducts in restriction categories	78
3.4.1 Coproducts	78
3.4.2 Inverse categories with restriction coproducts	80
3.5 Disjointness in an inverse category	81
3.5.1 Disjointness relations	81
3.5.2 Disjoint joins	87

3.5.3	Monoidal Tensors for disjointness	92
3.6	Inverse sum categories	106
3.6.1	Inverse sums	106
3.6.2	Inverse sum tensor	111
3.7	Completing a distributive inverse category	121
3.7.1	Distributive restriction categories	121
3.7.2	Distributive inverse categories	121
4	Quantum computation and circuits	123
4.1	Linear algebra	123
4.1.1	Basic definitions	123
4.1.2	Matrices	124
4.2	Basic quantum computation	126
4.2.1	Quantum bits	126
4.2.2	Quantum entanglement	127
4.2.3	Quantum gates	127
4.2.4	Measurement	128
4.2.5	Mixed states	129
4.2.6	Density matrix notation	129
4.2.7	Gates and density matrices	130
4.3	Quantum circuits	130
4.3.1	Contents of quantum circuits	130
4.3.2	Syntax of quantum circuits	135
4.3.3	Examples of quantum circuits	135
4.4	Extensions to quantum circuits	140
4.4.1	Renaming	140
4.4.2	Wire crossing	140
4.4.3	Scoped control	141
4.4.4	Circuit identities	142
4.5	An alternate description of quantum circuits	143
4.5.1	Base types	144
4.5.2	Types and Shapes	144
4.6	Quantum computation	145
5	Frobenius Algebras and Quantum Computation	151
6	$D[\omega]$ based \dagger categories	152
6.1	Introduction to synthesis	152
6.2	Algebraic background	152
6.2.1	Conjugate and norm	153
6.2.2	Denominator exponents	154
6.2.3	Residues	154
6.3	Exact synthesis of single qubit operators	157
6.3.1	Existence	159
6.3.2	T -Optimality	162
6.3.3	Uniqueness	162
6.3.4	The Matsumoto-Amano decomposition algorithm	166
6.3.5	A characterization of Clifford+ T on the Bloch sphere	167

6.3.6	Alternative normal forms	171
6.3.7	Matsumoto-Amano normal forms and $U(2)$	175
6.4	Exact synthesis of multi-qubit operators	178
6.4.1	Decomposition into two-level matrices	179
6.4.2	Main result	185
6.4.3	The no-ancilla case	187
6.4.4	Complexity	189
7	Conclusions and future work	191
	Bibliography	192

List of Tables

3.1	Structural maps for the tensor in $Inv(\mathbb{X})$	44
4.1	Gates, circuit notation and matrices	132
4.2	Syntactic elements of quantum circuit diagrams	136
6.1	Some operations on residues	155

List of Figures and Illustrations

4.1	Simple single gate circuit	131
4.2	Entangling two qubits	131
4.3	Controlled-Not of $ 1\rangle$ and $ 1\rangle$	131
4.4	Measure notation in quantum circuits	133
4.5	Examples of multi- qubit gates and measures	133
4.6	Other forms of control for gates	134
4.7	n qubits on one line	134
4.8	Swap and controlled-Z	134
4.9	Quantum teleportation	137
4.10	Circuit for the Deutsch-Jozsa algorithm	138
4.11	Circuit for the quantum Fourier transform	139
4.12	Circuit for the inverse quantum Fourier transform	140
4.13	Renaming of a qubit and its equivalent diagram	140
4.14	Bending	141
4.15	Scope of control	141
4.16	Extensions sample	141
4.17	Swap in control vs. exchange in control	142
4.18	Measure is not affected by control	142
4.19	Control is not affected by measure	142
4.20	Zero control is syntactic sugar	143
4.21	Scoped control is parallel control	143
4.22	Scoped control is serial control	143
4.23	Multiple control	144
4.24	Control scopes commute	144
6.1	The action of Matsumoto-Amano normal forms on k -parities. All matrices are written modulo the right action of the Clifford group, i.e., modulo a permutation of the columns.	164
6.2	Transitions of residue matrices in U2 when applying the Matsumoto-Amano algorithm	176

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
U of C	University of Calgary
\mathbb{N}	The set of natural numbers, i.e., $\{0, 1, 2, \dots\}$
\mathbb{Z}	The ring of integers numbers, i.e., $\{0, \pm 1, \pm 2, \dots\}$
\mathbb{C}	The field of complex numbers

Overview of thesis chapters

Introduction

This chapter will give a brief introduction to and an explanation of both reversible and quantum computing. How they are related will be discussed, along with a brief introduction to their categorical semantics. We will discuss the equivalence of reversible Turing machines and standard Turing machines.

Abstract Computability

This chapter will include a basic introduction to category theory. Specific areas introduced will include definitions of categories, natural transformations and functors. It will introduce limits and co-limits, focussing on products and co-products.

This chapter will start with an introduction to restriction categories and how Cartesian restriction categories can be used to model standard computing. We will introduce Turing categories and show how this can be used to create a Partial Combinatory Algebra.

Inverse categories and Reversible computing

An inverse category is a specific type of restriction category, in which each map has a partial inverse. An inverse category corresponds to a restriction category in the same way a groupoid corresponds to a category.

Inverse categories will be explored, along with some basic results regarding products and idempotent splitting. The inverse product will be introduced, along with the concept of a discrete inverse category and the relationship between a discrete inverse category and a Cartesian restriction category.

Next, we explore the inverse sum, disjointness and the disjoint join of maps in an inverse

category, allowing us to work with objects which behave like co-products in the inverse category. The interaction of the inverse sum and inverse product will be explained.

The chapter will define inverse Turing Categories.

The foregoing is based on two papers which are in preparation (joint work with R. Cockett).

To conclude the chapter, we will provide a detailed proof of the equivalence of reversible Turing machines and standard Turing machines and connect this to inverse categories.

Quantum Computation

The initial part of this chapter will introduce quantum circuits, following which, we will explore the semantics of quantum computing as described using \dagger -categories. This will include examples of “Toy” quantum semantics, specifically that of matrices over $D[\omega]$ (where $D = \mathbb{Z}[\frac{1}{2}]$).

Frobenius Algebras in \dagger -Categories

In this chapter, we highlight the connection between the model of reversible computing (inverse categories) introduced in this thesis, and that of quantum computing. Frobenius Algebras provide a way of describing the basis used in quantum computation for a specific model of quantum semantics.

$D[\omega]$ based \dagger categories and exact synthesis

We will revisit the example of a toy quantum semantics, that of the Clifford Group $+T$ over multiple qubits. As part of that, we will discuss the issue of gate synthesis, where an arbitrary quantum transform is to be expressed in terms of a set of base gates. An overview of the history of both approximate and exact synthesis will be given.

Then, the chapter will present an algorithm for exact synthesis of single-qubit transforms over the Clifford group, together with a normal form and characterization of these. This is based on a paper that is an extension of work done by Matsumoto and Amano (joint work with P. Selinger.)

Finally, we will present an algorithm for exact synthesis over the Clifford group of multi-qubit transforms and characterize those transforms that may be exactly synthesized. This is based on a paper published in the journal Physical Review A (joint work with P. Selinger).

Chapter 1

Introduction

Chapter 2

Abstract Computability

2.1 Categories

A category as a mathematical object can be defined in a variety of equivalent ways. As much of our work will involve the exploration of partial and reversible maps, their domains and ranges, we choose a definition that highlights the algebraic nature of these. Note that ranges are normally referred to as codomains in category theory and we will use the codomain terminology in this section.

Definition 2.1.1. A *category* \mathbb{A} is a collection of maps together with two functions, D and C , from \mathbb{A} to \mathbb{A} and a partial associative composition of maps (written by juxtaposing maps), such that:

$$[\mathbf{C.1}] \quad D(f)f \text{ is defined and equals } f,$$

$$[\mathbf{C.2}] \quad fC(f) \text{ is defined and equals } f,$$

$$[\mathbf{C.3}] \quad fg \text{ is defined iff } C(f) = D(g) \text{ and } D(fg) = D(f) \text{ and } C(fg) = C(g),$$

$$[\mathbf{C.4}] \quad (fg)h = f(gh) \text{ whenever either side is defined,}$$

$$[\mathbf{C.5}] \quad D(C(x)) = C(x), C(D(x)) = D(x) \text{ and } C, D \text{ are both idempotent.}$$

A more familiar definition, often used in introducing categories, is given next.

Definition 2.1.2. A *category* \mathbb{A} is a directed graph consisting of objects A_o and maps A_m . Each $f \in A_m$ has two associated objects in A_o , called the domain and codomain. When f has domain X and codomain Y we will write $f : X \rightarrow Y$. For $f, g \in A_m$, if $f : X \rightarrow Y$ and $g : Y \rightarrow Z$, there is a map called the *composite* of f and g , written fg such that $fg : X \rightarrow Z$.

For any $W \in A_o$ there is an *identity* map $1_W : W \rightarrow W$. Additionally, these two axioms must hold:

$$[\mathbf{C'}.1] \text{ for } f : X \rightarrow Y, 1_X f = f = f 1_Y,$$

$$[\mathbf{C'}.2] \text{ given } f : X \rightarrow Y, g : Y \rightarrow Z \text{ and } h : Z \rightarrow W, \text{ then } f(gh) = (fg)h.$$

Lemma 2.1.3. *A category as defined in Definition 2.1.1 is equivalent to a category as defined in Definition 2.1.2 and vice versa.*

Proof. Assume \mathbb{A} is as in Definition 2.1.1. Then set A_o to the collection of all $D(f)$ and $C(f)$. Set A_m to all the maps in \mathbb{A} . The domain of any map $f \in A_m$ is $D(f)$ and the codomain is $C(f)$. By $[\mathbf{C}.3]$, for $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ the composite fg is defined. The identity map of the object $D(f)$ is the map $D(f)$ and the identity map of the object $C(f)$ is $C(f)$. By $[\mathbf{C}.5]$, we see $[\mathbf{C'}.1]$ is satisfied. By $[\mathbf{C}.4]$, we see $[\mathbf{C'}.2]$ is satisfied. Therefore, \mathbb{A} satisfies Definition 2.1.2.

Conversely, assume \mathbb{Z} is as in Definition 2.1.2. Then, we already have the collection of maps, Z_m . For each $f : A \rightarrow B \in Z_m$, set $D(f) = 1_A$ and $C(f) = 1_B$. By the definition of the identity maps and $[\mathbf{C'}.1]$, we see $[\mathbf{C}.1]$, $[\mathbf{C}.2]$ and $[\mathbf{C}.5]$ are all satisfied. From the composition requirements on \mathbb{Z} and $[\mathbf{C'}.2]$, it follows that $[\mathbf{C}.4]$ is satisfied. For $[\mathbf{C}.3]$, assume fg is defined. Then for some $A, B, C \in Z_o$, $f : A \rightarrow B$ and $g : B \rightarrow C$. This gives us $1_B = C(f) = D(g)$, $1_A = D(fg) = D_f$ and $1_B = C(fg) = C(g)$. Next, assume we have $C(f) = D(g)$, $D(fg) = D(f)$ and $C(fg) = C(g)$. This tells us the codomain of f is some object B which is also the domain of g , hence we may form the composition fg which will have domain A , the domain of f and codomain C , the codomain of g . \square

As we have shown the two definitions are equivalent, it will be convenient to reference either definition and manner of referring to a category throughout this thesis. Essentially, we will use whichever definition seems the most appropriate to use at any point.

We may also consider the notion of containment between categories.

Definition 2.1.4. Given the categories \mathbb{C} and \mathbb{D} , we may say the following:

- (i) \mathbb{C} is a *sub-category* of \mathbb{D} when each object of \mathbb{C} is an object of \mathbb{D} and when each map of \mathbb{C} is a map of \mathbb{D} .
- (ii) \mathbb{C} is a *full sub-category* of \mathbb{D} when it is a sub-category and given A, B objects in \mathbb{C} and $f : A \rightarrow B$ in \mathbb{D} , then f is a map in \mathbb{C} .

2.1.1 Enrichment of categories

Definition 2.1.5. If \mathbb{X} is a category, then $\mathbb{X}(A, B)$ is called a *hom-collection* of \mathbb{X} and consists of all arrows f with $D(f) = A$ and $C(f) = B$.

In the case where the hom-objects of a category \mathbb{X} are all sets, we call them hom-sets. Additionally, we say \mathbb{X} is *enriched* in SETS. We may extend this to any mathematical structure, e.g., enriched in partial orders, enriched in groups, etc..

Specific types of enrichment may force a specific structure on a category. For example, if \mathbb{X} is enriched in sets of cardinality of 0 or 1, then \mathbb{X} must be a preorder.

2.1.2 Examples of categories

In this section, we will offer a few examples of categories. As Definition 2.1.2 tends to be a more succinct way to present the data of a category, this section will give the examples in terms of objects and maps rather than the “object-free” definition.

Categories based on SETS

There are three primary categories of interest to us where the objects are the collection of sets. The first is SETS, where the maps are given by all set functions. The second is PAR, where the maps are all partial maps. In each case, the standard definition of functions suffices to ensure identities, compositions and associativity are all satisfied. Domain and codomain are given by the domain and range respectively.

A third example, often of interest in quantum programming language semantics is REL:

Objects: Sets

Maps: Relations: $R : X \rightarrow Y$

Identity: $1_X = \{(x, x) | x \in X\}$

Composition: $RS = \{(x, z) | \exists y, (x, y) \in R \text{ and } (y, z) \in S\}$

Note that **REL** is enriched in posets, via set inclusion. **PAR** can be viewed as a subcategory of **REL**, with the same objects, but only allowing maps which are functions, i.e., if $(x, y), (x, y') \in R$, then $y = y'$. **PAR** is also enriched in posets, via the same inclusion ordering as in **REL**.

Matrix categories

Given a rig R (i.e., a ring minus negatives, e.g., the positive rationals), one may form the category **MAT** (R).

Objects: \mathbb{N}

Maps: $[r_{ij}] : n \rightarrow m$ where $[r_{ij}]$ is an $n \times m$ matrix over R

Identity: I_n

Composition: Matrix multiplication

Dual categories

Given a category \mathbb{C} , we may form the *dual* of \mathbb{C} , written \mathbb{C}^{op} as the following category:

Objects: The objects of \mathbb{C}

Maps: $f^{op} : B \rightarrow A$ in \mathbb{C}^{op} when $f : A \rightarrow B$ in \mathbb{C} .

Identity: The identity maps of \mathbb{C}

Composition: If $fg = h$ in \mathbb{C} , $g^{op}f^{op} = h^{op}$

2.1.3 Properties of maps

Many interesting properties of maps are generalizations of notions that have been found useful in considering sets and functions. We present a few of these in a tabular format, together with their categorical definition. Throughout the table, e, f, g are maps in a category C with $e : A \rightarrow A$ and $f, g : A \rightarrow B$.

Sets	Categorical Property	Definition
Injective	Monic	f is monic whenever $hf = kf$ means that $h = k$.
Surjective	Epic	The dual notion to monic, g is epic whenever $gh = gk$ means that $h = k$. A map that is both monic and epic is called <i>bijic</i> .
Left Inverse	Section	f is a section when there is a map f^* such that $ff^* = 1_A$. f is also referred to as the <i>left inverse</i> of f^* .
Right Inverse	Retraction	f is a retraction when there is a map f_* such that $f_*f = 1_B$. f is also referred to as the <i>right inverse</i> of f_* . A map that is both a section and a retraction is called an <i>isomorphism</i> .
Idempotent	Idempotent	An endomap e is idempotent whenever $ee = e$.

We state without proof a number of properties of maps.

Lemma 2.1.6. *In a category \mathbb{C} ,*

- (i) *If f, g are monic, then fg is monic.*
- (ii) *If fg is monic, then f is monic.*
- (iii) *f being a section means it is monic.*

(iv) f, g sections implies that fg is a section.

(v) fg a section means f is a section.

Lemma 2.1.7. *If $f : A \rightarrow B$ is both a section and a retraction, then $f^* = f_*$.*

Lemma 2.1.8. *f is an isomorphism if and only if it is an epic section.*

Note there are corresponding properties for epics and retractions, obtained by dualizing the statements of Lemma 2.1.6 and Lemma 2.1.8.

Suppose $f : A \rightarrow B$ is a retraction with left inverse $f_* : B \rightarrow A$. Note that ff_* is idempotent as $ff_*ff_* = f1_Bf_* = ff_*$. If we are given an idempotent e , we say e is *split* if there is a retraction f with $e = ff_*$.

In general, not all idempotents in a category will split. The following construction allows us to create a category based on the original one in which all idempotents do split.

Definition 2.1.9. Given a category \mathbb{C} we define $Split(\mathbb{C})$ as the following category:

Objects: (A, e) , where A is an object of \mathbb{C} , $e : A \rightarrow A$ and $e \in E$.

Maps: $f_{d,e} : (A, d) \rightarrow (B, e)$ is given by $f : A \rightarrow B$ in \mathbb{C} , where $f = dfe$.

Identity: The map $e_{e,e}$ for (A, e) .

Composition: Inherited from \mathbb{C} .

Lemma 2.1.10. *Given a category \mathbb{C} , then it is a full sub-category of $Split(\mathbb{C})$ and all idempotents split in $Split(\mathbb{C})$.*

Proof. We identify each object A in \mathbb{C} with the object $(A, 1)$ in $Split(\mathbb{C})$. The only maps between $(A, 1)$ and $(B, 1)$ in $Split(\mathbb{C})$ are the maps between A and B in \mathbb{C} , hence we have a full sub-category.

Suppose we have the map $d_{e,e} : (A, e) \rightarrow (A, e)$ with $dd = d$, i.e., it is idempotent in \mathbb{C} and $Split(\mathbb{C})$. In $Split(\mathbb{C})$, we have the map $d_{e,d} : (A, e) \rightarrow (A, d)$ and $d_{d,e} : (A, d) \rightarrow (A, e)$ where $d_{d,e}d_{e,d} = d_{d,d} = 1_{(A,d)}$ and $d_{e,d}d_{d,e} = d_{e,e}$, hence it is a splitting of the map $d_{e,e}$. \square

2.1.4 Limits and colimits in categories

We shall discuss only a few basic limits/colimits in categories. First we discuss initial and terminal objects.

Definition 2.1.11. An *initial object* in a category \mathbb{C} is an object which has exactly one map to each other object in the category. The dual notion is *terminal object* which has exactly one map from each other object in the category.

Lemma 2.1.12. Suppose I, J are initial objects in \mathbb{C} . Then there is a unique isomorphism $i : I \rightarrow J$.

Proof. First, note that by definition there is only one map from I to I — which must be the identity map. As I is initial there is a map $i : I \rightarrow J$. As J is initial there is a map $j : J \rightarrow I$. But this means $ij : I \rightarrow I = 1$ and $ji : J \rightarrow J = 1$ and hence i is the unique isomorphism from I to J . \square

Dually, we have the corresponding result of Lemma 2.1.12 for terminal objects — they are also unique up to a unique isomorphism.

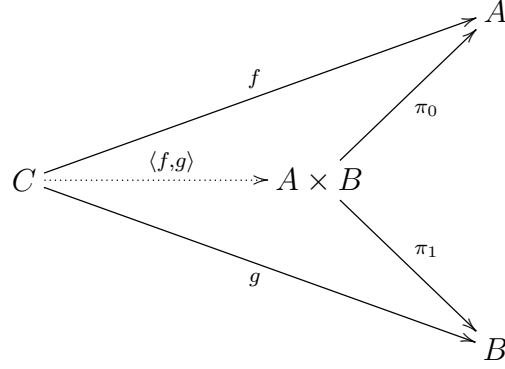
In categories, we normally designate the initial object by 0 and the terminal object by 1 .

We now turn to products and co-products.

Definition 2.1.13. Let A, B be objects of the category \mathbb{C} . Then the object $A \times B$ is a *product* of A and B when:

- There exist maps π_0, π_1 with $\pi_0 : A \times B \rightarrow A$, $\pi_1 : A \times B \rightarrow B$;
- Given an object C with maps $f : C \rightarrow A$ and $g : C \rightarrow B$ there exists an

unique map $\langle f, g \rangle$ such that the following diagram commutes:



2.1.5 Functors and natural transformations

Definition 2.1.14. A map $F : \mathbb{X} \rightarrow \mathbb{Y}$ between categories (as in Definition 2.1.1 is called a *functor*, provided it satisfies the following:

$$[\mathbf{F.1}] \quad F(D(f)) = D(F(f)) \text{ and } F(C(f)) = C(F(f));$$

$$[\mathbf{F.2}] \quad F(fg) = F(f)F(g);$$

Lemma 2.1.15. *The collection of categories and functors form the category CAT.*

Proof. **Objects:** Categories.

Maps: Functors.

Identity: The identity functor which takes a map to the same map.

Composition: $FG(x) = F(G(x))$ which is clearly associative.

□

We will often restrict ourselves to specific classes of functors which either *preserve* or *reflect* certain characteristics of the domain category or codomain category. To be more precise, we provide some definitions.

Definition 2.1.16. A *diagram* in a category is a collection of objects and maps between those objects which satisfy categorical composition rules. More precisely: Given a category \mathbb{C} , a diagram in a category \mathbb{C} of *shape* \mathbb{J} is a functor $D : \mathbb{J} \rightarrow \mathbb{C}$.

In practice, diagrams are pictorially represented by drawing the objects and the maps between them.

Definition 2.1.17. A *property* of a diagram D , written $P(D)$ is a logical relation expressed using the objects and maps of the diagram D .

Example 2.1.18. $P(f : A \rightarrow B) = \exists h : B \rightarrow A. hf = 1_A$ expresses that f is a retraction.

Definition 2.1.19. A functor F *preserves* the property P over maps f_i and objects A_j when $P(f_1, \dots, f_n, A_1, \dots, A_m)$ implies $P(F(f_1), \dots, F(f_n), F(A_1), \dots, F(A_m))$.

Definition 2.1.20. A functor F *reflects* the property P over maps f_i and objects A_j when $P(F(f_1), \dots, F(f_n), F(A_1), \dots, F(A_m))$ implies $P(f_1, \dots, f_n, A_1, \dots, A_m)$.

For example, all functors preserve the properties of being an idempotent or a retraction or section, but in general, not the property of being monic.

A functor $F : \mathbb{C} \rightarrow \mathbb{D}$ induces a map between hom-objects in \mathbb{C} and hom-objects in \mathbb{D} . For each object A, B in \mathbb{C} we have the map:

$$F_{AB} : \mathbb{C}(A, B) \rightarrow \mathbb{D}(F(A), F(B)).$$

Definition 2.1.21. Given a functor $F : \mathbb{C} \rightarrow \mathbb{D}$, we say:

- F is *faithful* when for all A, B , F_{AB} is an injective function;
- F is *full* when for all A, B , F_{AB} is a surjective function.

Definition 2.1.22. Given functors $F, G : \mathbb{X} \rightarrow \mathbb{Y}$, a *natural transformation* $\alpha : F \Rightarrow G$ is a collection of maps in \mathbb{Y} , $\alpha_X : F(X) \rightarrow G(X)$, indexed by the objects of \mathbb{X} such that for all $f : X_1 \rightarrow X_2$ in \mathbb{X} the following diagram in \mathbb{Y} commutes:

$$\begin{array}{ccc} F(X_1) & \xrightarrow{F(f)} & F(X_2) \\ \alpha_{X_1} \downarrow & & \downarrow \alpha_{X_2} \\ G(X_1) & \xrightarrow{G(f)} & G(X_2) \end{array}$$

2.2 Restriction categories

Restriction categories were introduced in [20] as a convenient axiomatization of partial maps.

Definition 2.2.1. A *restriction category* is a category \mathbb{X} together with a *restriction operator* on maps:

$$\frac{f : A \rightarrow B}{\overline{f} : A \rightarrow A}$$

where f is an map of \mathbb{X} and A, B are objects of \mathbb{X} , such that the following four *restriction identities* hold, whenever the compositions¹ are defined.

$$[\mathbf{R.1}] \quad \overline{f}f = f$$

$$[\mathbf{R.2}] \quad \overline{g}f = \overline{f}g$$

$$[\mathbf{R.3}] \quad \overline{\overline{f}g} = \overline{f}g$$

$$[\mathbf{R.4}] \quad f\overline{g} = \overline{f}gf$$

Definition 2.2.2. A *restriction functor* is a functor which preserves the restriction. That is, given a functor $F : \mathbb{X} \rightarrow \mathbb{Y}$ with \mathbb{X} and \mathbb{Y} restriction categories, F is a restriction functor if:

$$F(\overline{f}) = \overline{F(f)}.$$

Any map such that $r = \overline{r}$ is an idempotent, as $\overline{rr} = \overline{\overline{r}r} = \overline{r}$, and is called a *restriction idempotent*. All maps \overline{f} are restriction idempotents as $\overline{f} = \overline{\overline{f}}$. Below, we record some basic facts for restriction categories shown in [20] pp 4-5:

Lemma 2.2.3. *In a restriction category \mathbb{X} ,*

$$(i) \quad \overline{f} \text{ is idempotent};$$

$$(v) \quad \overline{f}\overline{g} = \overline{\overline{f}g};$$

$$(ii) \quad \overline{f}g = \overline{fg}\overline{f};$$

$$(vi) \quad f \text{ monic implies } \overline{f} = 1;$$

$$(iii) \quad \overline{f}g = \overline{f\overline{g}};$$

$$(vii) \quad f = \overline{g}f \implies \overline{g}\overline{f} = \overline{f}.$$

$$(iv) \quad \overline{\overline{f}} = \overline{f};$$

¹Note that composition is written in diagrammatic order throughout this paper.

A map $f : A \rightarrow B$ in a restriction category is said to be *total* when $\bar{f} = 1_A$. The total maps in a restriction category form a subcategory $Total(\mathbb{X}) \subseteq \mathbb{X}$.

An example of a restriction category is **PAR**, the category with objects sets and arrows the partial functions between sets. In **PAR**, the restriction of $f : A \rightarrow B$ is:

$$\bar{f}(x) = \begin{cases} x & \text{if } f(x) \text{ is defined,} \\ \uparrow & \text{if } f(x) \text{ is } \uparrow. \end{cases}$$

(The symbol \uparrow means that the function is undefined at that element). In **PAR**, the total maps correspond precisely to the functions that are defined on all elements of the domain.

2.2.1 Enrichment and meets

In any restriction category, there is a partial order on each hom-set, given by $f \leq g$ iff $\bar{f}g = f$, where $f, g : A \rightarrow B$.

Lemma 2.2.4. *In a restriction category \mathbb{X} :*

- (i) \leq as defined above is a partial order on each hom-set;
- (ii) $f \leq g \implies \bar{f} \leq \bar{g}$;
- (iii) $f \leq g \implies hf \leq hg$;
- (iv) $f \leq g \implies fh \leq gh$;
- (v) $f \leq 1 \iff f = \bar{f}$.

Proof.

- (i) With f, g, h parallel maps in \mathbb{X} , each of the requirements for a partial order is shown below:

Reflexivity: $\bar{f}f = f$ and therefore, $f \leq f$.

Anti-Symmetry: Given $\bar{f}g = f$ and $\bar{g}f = g$, it follows:

$$f = \bar{f}f = \overline{\bar{f}g}f = \bar{f}\bar{g}f = \bar{g}\bar{f}f = \bar{g}f = g.$$

Transitivity: Given $f \leq g$ and $g \leq h$,

$$\bar{f}h = \overline{\bar{f}g}h = \bar{f}\bar{g}h = \bar{f}g = f$$

showing that $f \leq h$.

(ii) The premise is that $\bar{f}g = f$. From this, $\bar{f}\bar{g} = \overline{\bar{f}g} = \bar{f}$, showing $\bar{f} \leq \bar{g}$.

(iii) $\bar{h}\bar{f}hg = h\bar{f}g = hf$ and therefore $hf \leq hg$.

(iv) $\bar{f}g = f$, this shows $\bar{f}hgh = \overline{\bar{f}g}hgh = \bar{f}\bar{g}hgh = \bar{f}gh = fh$ and therefore $fh \leq gh$.

(v) As $f \leq 1$ means precisely $\bar{f}1 = f$.

□

Lemma 2.2.4 on the preceding page shows that restriction categories are enriched in partial orders.

Definition 2.2.5. A restriction category has *meets* if there is an operation \cap on parallel maps:

$$\frac{A \xrightarrow{f} B}{A \xrightarrow{f \cap g} B}$$

such that $f \cap g \leq f, f \cap g \leq g, f \cap f = f, h(f \cap g) = hf \cap hg$.

Meets were introduced in [9]. The following are basic results on meets:

Lemma 2.2.6. *In a restriction category \mathbb{X} with meets, where f, g, h are maps in \mathbb{X} , the following are true:*

(i) $f \leq g$ and $f \leq h \iff f \leq g \cap h$;

- (ii) $f \cap g = g \cap f$;
- (iii) $\overline{f \cap 1} = f \cap 1$;
- (iv) $(f \cap g) \cap h = f \cap (g \cap h)$;
- (v) $r(f \cap g) = rf \cap g$ where $r = \bar{r}$ is a restriction idempotent;
- (vi) $(f \cap g)r = fr \cap g$ where $r = \bar{r}$ is a restriction idempotent;
- (vii) $\overline{f \cap g} \leq \bar{f}$ (and therefore $\overline{f \cap g} \leq \bar{g}$);
- (viii) $(f \cap 1)f = f \cap 1$;
- (ix) $e(e \cap 1) = e$ where e is idempotent.

Proof.

- (i) $f \leq g$ and $f \leq h$ means precisely $f = \bar{f}g$ and $f = \bar{f}h$. Therefore,

$$\bar{f}(g \cap h) = \bar{f}g \cap \bar{f}h = f \cap f = f$$

and so $f \leq g \cap h$. Conversely, given $f \leq g \cap h$, we have $f = \bar{f}(g \cap h) = \bar{f}g \cap \bar{f}h \leq \bar{f}g$. But $f \leq \bar{f}g$ means $f = \bar{f}\bar{f}g = \bar{f}g$ and therefore $f \leq g$. Similarly, $f \leq h$.

- (ii) From (i) on the previous page, as by definition, $f \cap g \leq g$ and $f \cap g \leq f$.
- (iii) $f \cap 1 = \overline{f \cap 1}(f \cap 1) = (\overline{f \cap 1}f) \cap (\overline{f \cap 1}) \leq \overline{f \cap 1}$ from which the result follows.
- (iv) By definition and transitivity, $(f \cap g) \cap h \leq f, g, h$ therefore by (i) on the preceding page $(f \cap g) \cap h \leq f \cap (g \cap h)$. Similarly, $f \cap (g \cap h) \leq (f \cap g) \cap h$ giving the equality.
- (v) Given $rf \cap g \leq rf$, calculate:

$$rf \cap g = \overline{rf \cap g}rf = \overline{r(f \cap g)}f = \overline{rrf \cap rgf} = \overline{r(f \cap g)}f = rf \cap gf = r(f \cap g).$$

(vi) Using the previous point with the restriction idempotent \overline{fr} ,

$$\begin{aligned} fr \cap g &= f\bar{r} \cap g = \overline{fr}f \cap g = \overline{fr}(f \cap g) = \overline{fr} \overline{f \cap g} f \\ &= \overline{f \cap g} \overline{fr} f = \overline{f \cap g} f\bar{r} = (f \cap g)r. \end{aligned}$$

(vii) For the first claim,

$$\overline{f \cap g} \bar{f} = \overline{\bar{f}(f \cap g)} = \overline{(\bar{f}f) \cap g} = \overline{f \cap g}.$$

The second claim then follows by (ii) on the previous page.

(viii) Given $f \cap 1 \leq f$:

$$f \cap 1 \leq f \iff \overline{f \cap 1}f = f \cap 1 \iff (f \cap 1)f = f \cap 1$$

where the last step is by item (iii) on the preceding page of this lemma.

(ix) As e is idempotent, $e(e \cap 1) = (ee \cap e) = e$.

□

2.2.2 Range categories

Corresponding to Definition 2.2.1, which axiomatizes the concept of a domain of definition, we now introduce range categories which serve to axiomatize the concept of the range of a function.

Definition 2.2.7. A restriction category \mathbb{X} is a *range category* when it has an operator on all maps

$$\frac{f : A \rightarrow B}{\hat{f} : B \rightarrow B}$$

where the operator satisfies the following:

$$[\mathbf{RR.1}] \quad \overline{\hat{f}} = \hat{f} \qquad [\mathbf{RR.2}] \quad f\hat{f} = f$$

$$[\mathbf{RR.3}] \quad \widehat{f\bar{g}} = \hat{f}\bar{g} \qquad [\mathbf{RR.4}] \quad \widehat{\hat{f}g} = \widehat{\hat{f}}\hat{g}$$

whenever the compositions are defined.

Lemma 2.2.8. *In a range category \mathbb{X} , the following hold:*

$$\begin{array}{ll}
(i) \quad \hat{g}\hat{f} = \hat{f}\hat{g}; & (v) \quad \hat{f}\hat{f} = \hat{f}; \\
(ii) \quad \overline{f}\hat{g} = \hat{g}\overline{f}; & (vi) \quad \hat{\hat{f}} = \hat{f}; \\
(iii) \quad \widehat{f\hat{g}} = \hat{f}\hat{g}; & (vii) \quad \widehat{\hat{f}} = \overline{f}; \\
(iv) \quad \hat{f} = 1 \text{ when } f \text{ is epic, hence} & (viii) \quad \widehat{\hat{g}\hat{f}\hat{g}} = \widehat{\hat{f}\hat{g}}; \\
\hat{1} = 1; & (ix) \quad \widehat{\hat{f}\hat{g}} = \hat{f}\hat{g}.
\end{array}$$

Proof. See, e.g., [11]. □

Lemma 2.2.9. *In a range category:*

$$\begin{array}{ll}
(i) \quad \overline{f}g \leq \overline{f}; & (iii) \quad f' \leq f \text{ implies } \overline{f'} \leq \overline{f}; \\
(ii) \quad \widehat{hf} \leq \hat{f}; & (iv) \quad f' \leq f \text{ implies } \hat{f'} \leq \hat{f}.
\end{array}$$

Proof.

- (i) This is immediate from Lemma 2.2.3[(ii)].
- (ii) Noting that $\widehat{hf}\hat{f} = \widehat{hf}\hat{f} = \widehat{hf}\hat{f} = \widehat{hf}$, we see $\widehat{hf} \leq \hat{f}$.
- (iii) $\overline{f'}\overline{f} = \overline{f'f} = \overline{f'}$, therefore $\overline{f'} \leq \overline{f}$.
- (iv) Calculating $\widehat{\hat{f'}}\hat{f} = \hat{f'}\hat{f} = \widehat{\hat{f'}}\hat{f} = \widehat{\hat{f'}}\hat{f} = \widehat{\hat{f'}}\hat{f} = \hat{f'}$, we see $\hat{f'} \leq \hat{f}$.

□

Remark 2.2.10. Note that unlike restrictions, a range is a *property* of a restriction category.

To see this, assume we have two ranges $\widehat{(-)}$ and $\widetilde{(-)}$. Then,

$$\hat{f} = \widehat{\widetilde{f}} = \hat{f}\widetilde{f} = \widetilde{\hat{f}} = \widetilde{\hat{f}} = \widetilde{\hat{f}} = \widetilde{f}.$$

Lemma 2.2.11. *An inverse category \mathbb{X} is a range category, where $\hat{f} = f^{(-1)}f = \overline{f^{(-1)}}$.*

Proof.

$$[\mathbf{RR.1}] \quad \widehat{f} = \overline{\overline{f^{(-1)}}} = \overline{f^{(-1)}} = \widehat{f};$$

$$[\mathbf{RR.2}] \quad f\widehat{f} = f\overline{f^{(-1)}} = f\overline{f^{(-1)}}f = \overline{f}f = f;$$

$$[\mathbf{RR.3}] \quad \widehat{f\overline{g}} = \overline{(f\overline{g})^{(-1)}} = \overline{\overline{g^{(-1)}}f^{(-1)}} = \overline{\overline{g}f^{(-1)}} = \overline{\overline{g}}\overline{f^{(-1)}} = \overline{f^{(-1)}}\overline{g} = \widehat{f\overline{g}};$$

$$[\mathbf{RR.4}] \quad \widehat{\widehat{f}g} = \overline{(f^{(-1)}g)^{(-1)}} = \overline{g^{(-1)}\overline{f^{(-1)}}^{(-1)}} = \overline{g^{(-1)}\overline{\overline{f^{(-1)}}}} = \overline{g^{(-1)}f^{(-1)}} = \overline{g^{(-1)}f^{(-1)}} = \overline{(fg)^{(-1)}} = \widehat{fg}$$

□

2.2.3 Partial monics, sections and isomorphisms

Partial isomorphisms play a central role in this paper and below we develop some their basic properties.

Definition 2.2.12. A map f in a restriction category \mathbb{X} is said:

- To be a *partial isomorphism* when there is a *partial inverse*, written $f^{(-1)}$ with $f f^{(-1)} = \overline{f}$ and $f^{(-1)} f = \overline{f^{(-1)}}$;
- To be a *partial monic* if $hf = kf \implies h\overline{f} = k\overline{f}$;
- To be a *partial section* if there exists an h such that $fh = \overline{f}$;
- To be a *restriction monic* if it is a section s with a retraction r such that $rs = \overline{rs}$.

Lemma 2.2.13. *In a restriction category:*

- (i) f, g partial monic implies fg is partial monic;
- (ii) f a partial section implies f is partial monic;
- (iii) f, g partial sections implies fg is a partial section;
- (iv) The partial inverse of f , when it exists, is unique;
- (v) If f, g have partial inverses and $f g$ exists, then $f g$ has a partial inverse;
- (vi) A restriction monic s is a partial isomorphism.

Proof.

(i) Suppose $hfg = kfg$. As g is partial monic, $hf\bar{g} = kf\bar{g}$. Therefore:

$$h\overline{f\bar{g}}f = k\overline{f\bar{g}}f \quad [\mathbf{R.4}]$$

$$h\overline{f\bar{g}}\bar{f} = k\overline{f\bar{g}}\bar{f} \quad f \text{ partial monic}$$

$$h\overline{f\bar{g}} = k\overline{f\bar{g}} \quad \text{Lemma 2.2.3, (ii)}$$

(ii) Suppose $gf = kf$. Then, $g\bar{f} = gfh = kfh = k\bar{f}$.

(iii) We have $fh = \bar{f}$ and $gh' = \bar{g}$. Therefore,

$$fgh'h = f\bar{g}h \quad g \text{ partial section}$$

$$= \overline{f\bar{g}}fh \quad [\mathbf{R.4}]$$

$$= \overline{f\bar{g}}\bar{f} \quad f \text{ partial section}$$

$$= \overline{f\bar{f}\bar{g}} \quad [\mathbf{R.2}]$$

$$= \overline{\bar{f}f\bar{g}} \quad [\mathbf{R.3}]$$

$$= \overline{f\bar{g}} \quad [\mathbf{R.1}]$$

(iv) Suppose both $f^{(-1)}$ and f^* are partial inverses of f . Then,

$$\begin{aligned} f^{(-1)} &= \overline{f^{(-1)}}f^{(-1)} = f^{(-1)}ff^{(-1)} = f^{(-1)}\bar{f} = f^{(-1)}ff^* = f^{(-1)}f\bar{f}^*f^* \\ &= \overline{f^{(-1)}\bar{f}^*f^*} = \overline{f^*f^{(-1)}}f^* = f^*f\overline{f^{(-1)}}f^* = f^*ff^{(-1)}ff^* = f^*ff^* = f^* \end{aligned}$$

(v) For $f : A \rightarrow B$, $g : B \rightarrow C$ with partial inverses $f^{(-1)}$ and $g^{(-1)}$ respectively, the partial inverse of fg is $g^{(-1)}f^{(-1)}$. Calculating $fgg^{(-1)}f^{(-1)}$ using all the restriction identities:

$$fgg^{(-1)}f^{(-1)} = f\bar{g}f^{(-1)} = \overline{f\bar{g}}ff^{(-1)} = \overline{f\bar{g}}\bar{f} = \overline{f\bar{f}\bar{g}} = \overline{\bar{f}f\bar{g}} = \overline{f\bar{g}}.$$

The calculation of $g^{(-1)}f^{(-1)}fg = \overline{g^{(-1)}f^{(-1)}}$ is similar.

(vi) The partial inverse of s is $\overline{rs}r$. First, note that $\overline{\overline{rs}r} = \overline{rs}\overline{r} = \overline{r}\overline{rs} = \overline{\overline{r}rs} = \overline{rs}$.

Then, it follows that $(\overline{rs}r)s = rs = \overline{rs} = \overline{\overline{rs}r}$ and $s(\overline{rs}r) = sr\overline{s} = \overline{s}$.

□

A restriction category in which every map is a partial isomorphism is called an *inverse category*.

An interesting property of inverse categories:

Lemma 2.2.14. *In an inverse category, all idempotents are restriction idempotents.*

Proof. Given an idempotent e ,

$$\overline{e} = ee^{(-1)} = eee^{(-1)} = e\overline{e} = \overline{e}e = \overline{e}e = e.$$

□

2.2.4 Split restriction categories

The split restriction category, $K_E(\mathbb{X})$ is defined as:

Objects: (A, e) , where A is an object of \mathbb{X} , $e : A \rightarrow A$ and $e \in E$.

Maps: $f : (A, d) \rightarrow (B, e)$ is given by $f : A \rightarrow B$ in \mathbb{X} , where $f = dfe$.

Identity: The map e for (A, e) .

Composition: inherited from \mathbb{X} .

This is the standard idempotent splitting construction, also known as the Karoubi envelope.

Note that for $f : (A, d) \rightarrow (B, e)$, by definition, in \mathbb{X} we have $f = dfe$, giving

$$df = d(dfe) = ddfe = dfe = f \quad \text{and} \quad fe = (dfe)e = dfee = dfe = f.$$

When \mathbb{X} is a restriction category, there is an immediate candidate for a restriction in $K_E(\mathbb{X})$.

If $f \in K_E(\mathbb{X})$ is e_1fe_2 in \mathbb{X} , then define \overline{f} as given by $e_1\overline{f}$ in \mathbb{X} . Note that for $f : (A, d) \rightarrow (B, e)$, in \mathbb{X} we have:

$$d\overline{f} = \overline{d}fd = \overline{f}d.$$

Proposition 2.2.15. *If \mathbb{X} is a restriction category and E is a set of idempotents, then the restriction as defined above makes $K_E(\mathbb{X})$ a restriction category.*

Proof. The restriction takes $f : (A, e_1) \rightarrow (B, e_2)$ to an endomorphism of (A, e_1) . The restriction is in $K_E(\mathbb{X})$ as

$$e_1(e_1\bar{f})e_1 = e_1\bar{f}e_1 = \overline{e_1\bar{f}}e_1e_1 = \overline{e_1\bar{f}}e_1 = e_1\bar{f}.$$

Checking the 4 restriction axioms:

$$[\mathbf{R.1}] \llbracket \bar{f}f \rrbracket = e_1\bar{f}f = e_1f = \llbracket f \rrbracket$$

$$[\mathbf{R.2}] \llbracket \bar{g}\bar{f} \rrbracket = e_1\bar{g}e_1\bar{f} = e_1e_1\bar{g}\bar{f} = e_1e_1\bar{f}\bar{g} = e_1\bar{f}e_1\bar{g} = \llbracket \bar{f}\bar{g} \rrbracket$$

$$[\mathbf{R.3}] \llbracket \bar{f}\bar{g} \rrbracket \equiv e_1\overline{e_1\bar{f}\bar{g}} = e_1\overline{e_1\bar{f}}\bar{g}e_1 = \overline{e_1\bar{f}}\bar{g}e_1 = e_1\overline{\bar{f}\bar{g}} = e_1\bar{f}\bar{g} = e_1e_1\bar{f}\bar{g} = e_1\bar{f}e_1\bar{g} = \llbracket \bar{f}\bar{g} \rrbracket$$

$$[\mathbf{R.4}] \llbracket \bar{f}\bar{g} \rrbracket = e_1f e_2\bar{g} = \overline{e_1f e_2\bar{g}}e_1f e_2 = \overline{e_1e_1f e_2\bar{g}}e_1f e_2 \\ = e_1\overline{e_1f e_2\bar{g}}e_1f e_2 = e_1\bar{f}\bar{g}e_1f e_2 = \llbracket \bar{f}\bar{g}f \rrbracket$$

□

Given this, provided all identity maps are in E , $K_E(\mathbb{X})$ is a restriction category with \mathbb{X} as a full sub-restriction category, via the embedding defined by taking an object A in \mathbb{X} to the object $(A, 1)$ in $K_E(\mathbb{X})$. Furthermore, the property of being an inverse category is preserved by splitting.

Lemma 2.2.16. *When \mathbb{X} is an inverse category, $K_E(\mathbb{X})$ is an inverse category.*

Proof. The inverse of $f : (A, e_1) \rightarrow (B, e_2)$ in $K_E(\mathbb{X})$ is $e_2f^{(-1)}e_1$ as

$$\llbracket ff^{(-1)} \rrbracket = e_1f e_2e_2f^{(-1)}e_1 = e_1e_1f e_2f^{(-1)}e_1 = e_1ff^{(-1)}e_1 = e_1e_1\bar{f}e_1 = e_1\bar{f} = \llbracket \bar{f} \rrbracket$$

and

$$\llbracket f^{(-1)}f \rrbracket = e_2f^{(-1)}e_1e_1f e_2 = e_2f^{(-1)}e_1f e_2e_2 = e_2f^{(-1)}f e_2 \\ = e_2e_2\overline{f^{(-1)}}e_2 = e_2\overline{f^{(-1)}} = \llbracket \overline{f^{(-1)}} \rrbracket$$

□

Proposition 2.2.17. *In a restriction category \mathbb{X} , with meets, let R be the set of restriction idempotents. Then, $K(\mathbb{X}) \cong K_R(\mathbb{X})$ (where $K(\mathbb{X})$ is the split of \mathbb{X} over all idempotents). Furthermore, $K_R(\mathbb{X})$ has meets.*

Proof. The proof below first shows the equivalence of the two categories, then addresses the claim that $K_R(\mathbb{X})$ has meets.

For equivalence, we require two functors,

$$U : K_R(\mathbb{X}) \rightarrow K(\mathbb{X}) \text{ and } V : K(\mathbb{X}) \rightarrow K_R(\mathbb{X}),$$

with:

$$UV \cong I_{K_R(\mathbb{X})} \tag{2.1}$$

$$VU \cong I_{K(\mathbb{X})}. \tag{2.2}$$

U is the standard inclusion functor. V will take the object (A, e) to $(A, e \cap 1)$ and the map $f : (A, e_1) \rightarrow (B, e_2)$ to $(e_1 \cap 1)f$.

V is a functor as:

Well Defined: If $f : (A, e_1) \rightarrow (B, e_2)$, then $(e_1 \cap 1)f$ is a map in \mathbb{X} from A to B and

$$(e_1 \cap 1)(e_1 \cap 1)f(e_2 \cap 1) = (e_1 \cap 1)(fe_2 \cap f) = (e_1 \cap 1)(f \cap f) = (e_1 \cap 1)f,$$

therefore, $V(f) : V((A, e_1)) \rightarrow V((B, e_2))$.

Identities: $V(e) = (e \cap 1)e = e \cap 1$ by lemma 2.2.6 on page 16.

Composition: $V(f)V(g) = (e_1 \cap 1)f(e_2 \cap 1)g = (e_1 \cap 1)fe_2(e_2 \cap 1)g = (e_1 \cap 1)f(e_2 \cap e_2)g = (e_1 \cap 1)fe_2g = (e_1 \cap 1)fg = V(fg)$.

Recalling from Lemma 2.2.6 on page 16, $(e \cap 1)$ is a restriction idempotent. Using this fact, the commutativity of restriction idempotents and the general idempotent identities from 2.2.6 on page 16, the composite functor UV is the identity on $K_r(\mathbb{X})$ as when e is a restriction idempotent, $e = e(e \cap 1) = (e \cap 1)e = (e \cap 1)$.

For the other direction, note that for a particular idempotent $e : A \rightarrow A$, this gives the maps $e : (A, e) \rightarrow (A, e \cap 1)$ and $e \cap 1 : (A, e \cap 1) \rightarrow (A, e)$, again by [2.2.6 on page 16](#). These maps give the natural isomorphism between I and VU as

$$\begin{array}{ccc} (A, e) & \xrightarrow{e} & (A, e \cap 1) \\ & \searrow e & \downarrow e \cap 1 \\ & & (A, e) \end{array} \quad \text{and} \quad \begin{array}{ccc} (A, e \cap 1) & \xrightarrow{e \cap 1} & (A, e) \\ & \searrow e \cap 1 & \downarrow e \\ & & (A, e \cap 1) \end{array}$$

both commute. Therefore, $UV = I$ and $VU \cong I$, giving an equivalence of the categories.

For the rest of this proof, the bolded functions, e.g., \mathbf{f} are in $K_R(\mathbb{X})$. Italic functions, e.g., f are in \mathbb{X} .

To show that $K_R(\mathbb{X})$ has meets, designate the meet in $K_R(\mathbb{X})$ as \cap_K and define $\mathbf{f} \cap_K \mathbf{g}$ as the map given by the \mathbb{X} map $f \cap g$, where $\mathbf{f}, \mathbf{g} : (A, d) \rightarrow (B, e)$ in $K_R(\mathbb{X})$ and $f, g : A \rightarrow B$ in \mathbb{X} . This is a map in $K_R(\mathbb{X})$ as $d(f \cap g)e = (df \cap dg)e = (f \cap g)e = (fe \cap g) = f \cap g$ where the penultimate equality is by [2.2.6 on page 16](#). By definition $\overline{\mathbf{f} \cap_K \mathbf{g}}$ is $\overline{df \cap dg}$.

It is necessary to show \cap_K satisfies the four meet properties.

- $\mathbf{f} \cap_K \mathbf{g} \leq \mathbf{f}$: We need to show $\overline{\mathbf{f} \cap_K \mathbf{g}} \mathbf{f} = \mathbf{f} \cap_K \mathbf{g}$. Calculating now in \mathbb{X} :

$$\begin{aligned} \overline{df \cap dg} f &= \overline{d(f \cap g)} df \\ &= \overline{df \cap dg} df \\ &= \overline{f \cap g} f \\ &= f \cap g \end{aligned}$$

which is the definition of $\mathbf{f} \cap_K \mathbf{g}$.

- $\mathbf{f} \cap_K \mathbf{g} \leq \mathbf{g}$: Similarly and once again calculating in \mathbb{X} ,

$$\begin{aligned} \overline{df \cap dg} g &= \overline{d(f \cap g)} dg \\ &= \overline{df \cap dg} dg \\ &= \overline{f \cap g} g \\ &= f \cap g \end{aligned}$$

which is the definition of $\mathbf{f} \cap_K \mathbf{g}$.

- $\mathbf{f} \cap_K \mathbf{f} = \mathbf{f}$: From the definition, this is $f \cap f = f$ which is just \mathbf{f} .
- $\mathbf{h}(\mathbf{f} \cap_K \mathbf{g}) = \mathbf{hf} \cap_K \mathbf{hg}$: From the definition, this is given in \mathbb{X} by $h(f \cap g) = hf \cap hg$ which in $K_R(\mathbb{X})$ is $\mathbf{hf} \cap_K \mathbf{hg}$.

□

2.2.5 Partial Map Categories

In [20], it is shown that split restriction categories are equivalent to *partial map categories*.

The main definitions and results related to partial map categories are given below.

Definition 2.2.18. A collection \mathcal{M} of monics is a *stable system of monics* when it includes all isomorphisms, is closed under composition and is pullback stable.

Stable in this definition means that if $m : A \rightarrow B$ is in \mathcal{M} , then for arbitrary b with codomain B , the pullback

$$\begin{array}{ccc} A' & \xrightarrow{a} & A \\ m' \downarrow & & \downarrow m \\ B' & \xrightarrow{b} & B \end{array}$$

exists and $m' \in \mathcal{M}$. A category that has a stable system of monics is referred to as an \mathcal{M} -category.

Lemma 2.2.19. If $nm \in \mathcal{M}$, a stable system of monics, and m is monic, then $n \in \mathcal{M}$.

Proof. The commutative square

$$\begin{array}{ccc} A & \xrightarrow{1} & A \\ n \downarrow & & \downarrow nm \\ A' & \xrightarrow{m} & B \end{array}$$

is a pullback.

□

Given a category \mathbb{C} and a stable system of monics, the *partial map category*, $\text{Par}(\mathbb{C}, \mathcal{M})$ is:

Objects: $A \in \mathbb{C}$

Equivalence Classes of Maps: $(m, f) : A \rightarrow B$ with $m : A' \rightarrow A$ is in \mathcal{M} and $f : A' \rightarrow B$

is a map in \mathbb{C} . i.e.,
$$\begin{array}{ccc} & A' & \\ m \swarrow & & \searrow f \\ A & & B \end{array} .$$

Identity: $1_A, 1_A : A \rightarrow A$

Composition: via a pullback, $(m, f)(m', g) = (m''m, f'g)$ where

$$\begin{array}{ccccc} & & A'' & & \\ & m'' \swarrow & & \searrow f' & \\ & A' & \text{(pb)} & B' & \\ m \swarrow & & & & \searrow g \\ A & & f \searrow & B & \swarrow m' \end{array}$$

Restriction: $\overline{(m, f)} = (m, m)$

For the maps, $(m, f) \sim (m', f')$ when there is an isomorphism $\gamma : A'' \rightarrow A'$ such that $\gamma m' = m$ and $\gamma f' = f$.

In [21], it is shown that:

Theorem 2.2.20 (Cockett-Lack). *Every restriction category is a full subcategory of a partial map category.*

2.2.6 Restriction products and Cartesian restriction categories

Restriction categories have analogues of products and terminal objects.

Definition 2.2.21. In a restriction category \mathbb{X} a *restriction product* of two objects X, Y is an object $X \times Y$ equipped with *total* projections $\pi_0 : X \times Y \rightarrow X, \pi_1 : X \times Y \rightarrow Y$ where:

$\forall f : Z \rightarrow X, g : Z \rightarrow Y, \exists$ a unique $\langle f, g \rangle : Z \rightarrow X \times Y$ such that

- $\langle f, g \rangle \pi_0 \leq f$,
- $\langle f, g \rangle \pi_1 \leq g$ and

- $\overline{\langle f, g \rangle} = \bar{f} \bar{g} (= \bar{g} \bar{f})$.

Definition 2.2.22. In a restriction category \mathbb{X} a *restriction terminal object* is an object \top such that $\forall X$, there is a unique total map $!_X : X \rightarrow \top$ and the diagram

$$\begin{array}{ccccc} X & \xrightarrow{\bar{f}} & X & \xrightarrow{!_X} & \top \\ \downarrow f & & & \nearrow !_Y & \\ Y & & & & \end{array}$$

commutes. That is, $f !_Y = \bar{f} !_X$. Note this implies that a restriction terminal object is unique up to a unique isomorphism.

Definition 2.2.23. A restriction category \mathbb{X} is *Cartesian* if it has all restriction products and a restriction terminal object.

Definition 2.2.24. An object A in a Cartesian restriction category is *discrete* when the diagonal map,

$$\Delta : A \rightarrow A \times A$$

is a partial isomorphism.

A Cartesian restriction category is *discrete* when every object is discrete.

Theorem 2.2.25. A Cartesian restriction category \mathbb{X} is discrete if and only if it has meets.

Proof. If \mathbb{X} has meets, then

$$\Delta(\pi_0 \cap \pi_1) = \Delta\pi_0 \cap \Delta\pi_1 = 1 \cap 1 = 1$$

and as $\langle \pi_0, \pi_1 \rangle$ is identity,

$$\begin{aligned} \overline{\pi_0 \cap \pi_1} &= \overline{\pi_0 \cap \pi_1} \langle \pi_0, \pi_1 \rangle \\ &= \langle \overline{\pi_0 \cap \pi_1} \pi_0, \overline{\pi_0 \cap \pi_1} \pi_1 \rangle \\ &= \langle \pi_0 \cap \pi_1, \pi_0 \cap \pi_1 \rangle \\ &= (\pi_0 \cap \pi_1) \Delta \end{aligned}$$

and therefore, $\pi_0 \cap \pi_1$ is $\Delta^{(-1)}$.

For the other direction, set $f \cap g = \langle f, g \rangle \Delta^{(-1)}$. By the definition of the restriction product:

$$f \cap g = \langle f, g \rangle \Delta^{(-1)} = \langle f, g \rangle \Delta^{(-1)} \Delta \pi_0 = \langle f, g \rangle \overline{\Delta^{(-1)}} \pi_0 \leq \langle f, g \rangle \pi_0 \leq f$$

Similarly, substituting π_1 for π_0 above, this gives $f \cap g \leq g$. For the left distributive law,

$$h(f \cap g) = h \langle f, g \rangle \Delta^{(-1)} = \langle hf, hg \rangle \Delta^{(-1)} = hf \cap hg$$

and finally an intersection of a map with itself is

$$f \cap f = \langle f, f \rangle \Delta^{(-1)} = (f \Delta) \Delta^{(-1)} = f \overline{\Delta} = f$$

as Δ is total. This shows that \cap as defined above is a meet for the Cartesian restriction category \mathbb{X} .

□

We shall refer to a Cartesian restriction category in which every object is discrete as simply a discrete restriction category.

2.2.7 Graphic Categories

In a Cartesian restriction category, a map $A \xrightarrow{f} B$ is called *graphic* when the maps

$$A \xrightarrow{\langle f, 1 \rangle} B \times A \quad \text{and} \quad A \xrightarrow{\langle \bar{f}, 1 \rangle} A \times A$$

have partial inverses. A Cartesian restriction category is *graphic* when all of its maps are graphic.

Lemma 2.2.26. *In a Cartesian restriction category:*

- (i) *Graphic maps are closed under composition;*
- (ii) *Graphic maps are closed under the restriction;*

(iii) An object is discrete if and only if its identity map is graphic.

Proof.

(i) To show closure, it is necessary to show that $\langle fg, 1 \rangle$ has a partial inverse. By Lemma 2.2.13 on page 20, the uniqueness of the partial inverse gives

$$(\langle f, 1 \rangle; \langle g, 1 \rangle \times 1)^{(-1)} = \langle g, 1 \rangle^{(-1)} \times 1; \langle f, 1 \rangle^{(-1)}.$$

By the definition of the restriction product, $\overline{\langle fg, 1 \rangle} = \overline{fg}$. Additionally, a straightforward calculation shows that $\overline{\langle f, 1 \rangle; \langle g, 1 \rangle \times 1} = \overline{\langle f \langle g, 1 \rangle, 1 \rangle} = \overline{f; \langle g, 1 \rangle} = \overline{\langle f; g, f \rangle} = \overline{fgf} = \overline{fg}$ where the last equality is by [R.2], [R.3] and finally [R.1].

Consider the diagram

$$\begin{array}{ccccc} A & \xrightarrow{\langle f, 1 \rangle} & B \times A & \xrightarrow{\langle g, 1 \rangle \times 1} & C \times B \times A \\ & \searrow \langle fg, 1 \rangle & & & \uparrow 1 \times \langle f, 1 \rangle \\ & & & & C \times A \end{array}$$

From this:

$$\begin{aligned} \langle fg, 1 \rangle (1 \times \langle f, 1 \rangle) (\langle g, 1 \rangle^{(-1)} \times 1) \langle f, 1 \rangle^{(-1)} &= \langle f, 1 \rangle (\langle g, 1 \rangle \times 1) (\langle g, 1 \rangle^{(-1)} \times 1) \langle f, 1 \rangle^{(-1)} \\ &= \langle f, 1 \rangle (\overline{g \times 1}) \langle f, 1 \rangle^{(-1)} \\ &= \overline{\langle f, 1 \rangle (g \times 1) \langle f, 1 \rangle \langle f, 1 \rangle^{(-1)}} \\ &= \overline{\langle f, 1 \rangle (g \times 1) \langle f, 1 \rangle} \\ &= \overline{\langle f, 1 \rangle \langle f, 1 \rangle (g \times 1)} \\ &= \overline{\langle f, 1 \rangle (g \times 1)} \\ &= \overline{\langle fg, 1 \rangle} (= \overline{fg}) \end{aligned}$$

showing that $1 \times \langle f, 1 \rangle (\langle g, 1 \rangle^{(-1)} \times 1) \langle f, 1 \rangle^{(-1)}$ is a right inverse for $\langle fg, 1 \rangle$.

For the other direction, note that in general $hk^{(-1)} = k^{(-1)}h^{(-1)}$ and that we have $\langle fg, 1 \rangle = \langle f, 1 \rangle (\langle g, 1 \rangle \times 1) (1 \times \langle f, 1 \rangle^{(-1)})$, thus $(1 \times \langle f, 1 \rangle) (\langle g, 1 \rangle^{(-1)} \times 1) \langle f, 1 \rangle^{(-1)}$ will also be a left inverse and $\langle fg, 1 \rangle$ is a restriction isomorphism.

- (ii) This follows from the definition of graphic and that $\overline{\langle f, 1 \rangle} = \bar{f} = \overline{\bar{f}}$.
- (iii) Given a discrete object A , the map 1_A is graphic as $\langle 1_A, 1 \rangle = \Delta$ and therefore $\langle 1, 1 \rangle^{(-1)} = \Delta^{(-1)}$. Conversely, if $\langle 1_A, 1 \rangle$ has an inverse, then $\Delta = \langle 1_A, 1 \rangle$ has that same inverse and therefore the object is discrete.

□

Lemma 2.2.27. *A discrete restriction category is precisely a graphic Cartesian restriction category.*

Proof. The requirement is that $\langle f, 1 \rangle$ (and $\langle \bar{f}, 1 \rangle$) each have partial inverses. For $\langle f, 1 \rangle$, the inverse is $\overline{(1 \times f) \Delta^{(-1)}} \pi_1$.

To show this, calculate the two compositions. First,

$$\langle f, 1 \rangle \overline{(1 \times f) \Delta^{(-1)}} \pi_1 = \overline{\langle f, f \rangle \Delta^{(-1)}} \langle f, 1 \rangle \pi_1 = \overline{f \Delta \Delta^{(-1)}} \langle f, 1 \rangle \pi_1 = \bar{f} \langle f, 1 \rangle \pi_1 = \bar{f}.$$

The other direction is:

$$\begin{aligned} \overline{(1 \times f) \Delta^{(-1)}} \pi_1 \langle f, 1 \rangle &= \langle \overline{(1 \times f) \Delta^{(-1)}} \pi_1 f, \overline{(1 \times f) \Delta^{(-1)}} \pi_1 \rangle \\ &= \langle \overline{(1 \times f) \Delta^{(-1)}} (1 \times f) \pi_1, \overline{(1 \times f) \Delta^{(-1)}} \pi_1 \rangle \\ &= \langle \overline{(1 \times f) \Delta^{(-1)}} \pi_1, \overline{(1 \times f) \Delta^{(-1)}} \pi_1 \rangle \\ &= \langle \overline{(1 \times f) \Delta^{(-1)}} \pi_0, \overline{(1 \times f) \Delta^{(-1)}} \pi_1 \rangle \\ &= \langle \overline{(1 \times f) \Delta^{(-1)}} (1 \times f) \pi_0, \overline{(1 \times f) \Delta^{(-1)}} \pi_1 \rangle \\ &= \langle \overline{(1 \times f) \Delta^{(-1)}} \pi_0, \overline{(1 \times f) \Delta^{(-1)}} \pi_1 \rangle \\ &= \overline{(1 \times f) \Delta^{(-1)}} \langle \pi_0, \pi_1 \rangle \\ &= \overline{(1 \times f) \Delta^{(-1)}} \end{aligned}$$

The one tricky step is to realize

$$\begin{aligned}
\overline{\Delta^{(-1)}}\pi_1 &= \Delta^{(-1)}\Delta\pi_1 \\
&= \Delta^{(-1)} \\
&= \Delta^{(-1)}\Delta\pi_0 \\
&= \overline{\Delta^{(-1)}}\pi_0
\end{aligned}$$

For $\langle \bar{f}, 1 \rangle$, the inverse is $\overline{(1 \times \bar{f})\Delta^{(-1)}}\pi_1$. Similarly to above,

$$\langle \bar{f}, 1 \rangle \overline{1 \times \bar{f}\Delta^{(-1)}}\pi_1 = \overline{\langle \bar{f}, \bar{f} \rangle \Delta^{(-1)}}\langle \bar{f}, 1 \rangle \pi_1 = \overline{\bar{f}\Delta\Delta^{(-1)}}\langle \bar{f}, 1 \rangle \pi_1 = \bar{\bar{f}}\langle \bar{f}, 1 \rangle \pi_1 = \bar{f}.$$

The other direction follows the same pattern as for $\langle f, 1 \rangle$. □

Chapter 3

Inverse categories

3.1 Inverse products

Our goal is now to add “products”, to an inverse category. Because an inverse category that has a restriction product is a restriction preorder, what is meant by “product” must be specialized for the inverse setting. These we call *inverse products*, which are defined in sub-section [sub-section 3.1.2 on page 35](#) below.

Inverse products are given by a tensor product which supports a diagonal, but lack projections. The diagonal map is required to give a natural Frobenius structure to each object.

3.1.1 Inverse categories with restriction products

We start by showing than an inverse category with restriction products is a restriction preorder. Thus simply using restriction products provides a notion which is too narrow.

Definition 3.1.1. Two parallel maps $f, g : A \rightarrow B$ in a restriction category are *compatible*, written as $f \smile g$, when $\bar{f}g = \bar{g}f$.

Definition 3.1.2. A restriction category \mathbb{X} is a *restriction preorder* when all parallel pairs of maps are compatible.

Lemma 3.1.3. *Given an inverse category \mathbb{X} , if it has restriction products, it is a restriction preorder. That is,*

$$A \begin{array}{c} \xrightarrow{f} \\ \xrightarrow{g} \end{array} B \implies f \smile g.$$

Proof. Notice,

$$\begin{aligned}
\pi_1^{(-1)} &= \Delta \pi_1 \pi_1^{(-1)} \\
&= \Delta \overline{\pi_1} \\
&= \Delta.
\end{aligned}$$

This gives $\overline{\pi_1^{(-1)}} = 1$ and therefore π_1 (and similarly, π_0) is an isomorphism.

Starting with the product map $\langle f, g \rangle$,

$$\begin{aligned}
&\overline{\langle f, g \rangle} = \langle f, g \rangle \\
&\overline{\langle f, g \rangle \pi_1 \pi_1^{(-1)}} = \overline{\langle f, g \rangle \pi_0 \pi_0^{(-1)}} \\
&\overline{\overline{f} g \pi_1^{(-1)}} = \overline{\overline{g} f \pi_0^{(-1)}} \\
&\overline{\overline{f} g \Delta} = \overline{\overline{g} f \Delta} \\
&\overline{\overline{f} g} = \overline{\overline{g} f}
\end{aligned}$$

which shows that f and g are compatible. □

Corollary 3.1.4. *\mathbb{X} is an Cartesian inverse category if and only if $Total(K_r(\mathbb{X}))$ is a meet preorder.*

Proof. $Total(\mathbb{X})$, the subcategory of total maps on \mathbb{X} , has products and therefore every pair of parallel maps is compatible. However, total compatible maps are simply equal, therefore there is at most one map between any two objects. Hence, it is a preorder with the meet being the product.

Similarly, from [20] and [22], $Total(K_r(\mathbb{X}))$ is an inverse category and has products and is therefore also a meet preorder.

When $Total(K_r(\mathbb{X}))$ is a meet preorder, define the product as the meet of the maps and the terminal object as the supremum of all maps. □

Corollary 3.1.5. *Every Cartesian inverse category is a full subcategory of a partial map category of a meet semi-lattice.*

3.1.2 Inverse products

An *inverse product* on a restriction category \mathbb{X} is given by a tensor \otimes together with a natural “Frobenius” diagonal map, Δ . The data for the tensor is:

$$- \otimes - : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{X} \quad (\text{a restriction functor})$$

$$1 : \mathbf{1} \rightarrow \mathbb{X}$$

$$u_{\otimes}^l : 1 \otimes A \rightarrow A$$

$$u_{\otimes}^r : A \otimes 1 \rightarrow A$$

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

$$c_{\otimes} : A \otimes B \rightarrow B \otimes A$$

where $u_{\otimes}^l, u_{\otimes}^r, a_{\otimes}, c_{\otimes}$ are all natural isomorphisms and the standard symmetric monoidal equations and coherence diagrams hold (see, e.g., [15]). Note that as all the coherence maps are isomorphisms, they are total. Additionally, we define the map $ex_{\otimes} : (A \otimes B) \otimes (C \otimes D) \rightarrow (A \otimes C) \otimes (B \otimes D)$

$$ex_{\otimes} = a_{\otimes}(1 \otimes a_{\otimes}^{(-1)})(1 \otimes (c_{\otimes} \otimes 1))(1 \otimes a_{\otimes})a_{\otimes}^{(-1)}.$$

The diagonal map $\Delta_A : A \rightarrow A \otimes A$ must be total and must satisfy the following:

$$\begin{array}{ccc} A & \xrightarrow{\Delta} & A \otimes A \\ & \searrow \Delta & \downarrow c_{\otimes} \\ & & A \otimes A \end{array}$$

Co-commutative

$$\begin{array}{ccc}
A & \xrightarrow{\Delta} & A \otimes A \\
\Delta \downarrow & & \downarrow 1 \otimes \Delta \\
A \otimes A & \xrightarrow{\Delta \otimes 1} & (A \otimes A) \otimes A \\
& \searrow \Delta \otimes 1 & \nearrow a_{\otimes} \\
& (A \otimes A) \otimes A &
\end{array}$$

Co-associative

$$\begin{array}{ccc}
A \otimes B & \xrightarrow{\Delta \otimes \Delta} & (A \otimes A) \otimes (B \otimes B) \\
\Delta \downarrow & & \downarrow ex_{\otimes} \\
(A \otimes B) \otimes (A \otimes B) & \xlongequal{\quad} & (A \otimes B) \otimes (A \otimes B)
\end{array}$$

Exchange

$$\begin{array}{ccccc}
A \otimes A & \xrightarrow{(\Delta \otimes 1)a_{\otimes}} & A \otimes (A \otimes A) & & \\
\downarrow (1 \otimes \Delta)a_{\otimes}^{(-1)} & \searrow \Delta^{(-1)} & \downarrow 1 \otimes \Delta^{(-1)} & & \\
(A \otimes A) \otimes A & \xrightarrow{\Delta^{(-1)} \otimes 1} & A \otimes A & & \\
& \nearrow \Delta & & &
\end{array}$$

Frobenius

Thus, Δ is a co-commutative, coassociative map which together with $\Delta^{(-1)}$ forms a Frobenius algebra.

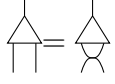
Remark 3.1.6. Note also, co-commutativity implies that $c_{\otimes} \Delta^{(-1)} = \Delta^{(-1)}$. One can see this as:

$$\begin{aligned}
\Delta(c_{\otimes} \Delta^{(-1)}) &= (\Delta c_{\otimes}) \Delta^{(-1)} = \Delta \Delta^{(-1)} = \overline{\Delta} \text{ and} \\
(c_{\otimes} \Delta^{(-1)}) \Delta &= (c_{\otimes} \Delta^{(-1)}) (\Delta c_{\otimes}) = \overline{c_{\otimes} \Delta^{(-1)}}.
\end{aligned}$$

But this means that both $\Delta^{(-1)}$ and $c_{\otimes}\Delta^{(-1)}$ are partial inverses for Δ and are therefore equal.

Similarly, one can show that $(\Delta^{(-1)} \otimes 1)\Delta^{(-1)} = a_{\otimes}(\Delta^{(-1)} \otimes 1)\Delta^{(-1)}$.

Diagrammatic language



Inverse products are extra structure on an inverse category, rather than a property. A concrete category showing this is given in the following example.

Example 3.1.7 (Showing that inverse product is additional structure.).

Any discrete category (i.e., a category with only the identity arrows) is a trivial inverse category. To create an inverse product on the category, add a commutative, associative, idempotent multiplication, with a unit, on the objects.

Label the four objects of \mathbb{D} as a, b, c and d . Then, define two different inverse product tensors by:

\otimes	a	b	c	d
a	a	a	a	a
b	a	b	b	b
c	a	b	c	c
d	a	b	c	d

\otimes	a	b	c	d
a	a	a	a	a
b	a	b	a	b
c	a	a	c	c
d	a	b	c	d

The fact that these operations are idempotent(commutative and associative) implies there is a trivial Frobenius structure.

3.1.3 Discrete inverse categories

An inverse category with inverse products is a *discrete inverse category*. This paper will now present some properties of discrete inverse categories. These properties will be used later when describing a functor that lifts the inverse category to a Cartesian restriction category.

Lemma 3.1.8. *In a discrete inverse category \mathbb{X} with the tensor \otimes and Δ defined as above, where $e = \bar{e}$ is a restriction idempotent and f, g, h are arrows in \mathbb{X} , the following are true:*

- (i) $e = \Delta(e \otimes 1)\Delta^{(-1)}$.
- (ii) $e\Delta(f \otimes g) = \Delta(ef \otimes g)$ (and $= \Delta(f \otimes eg)$ and $= \Delta(ef \otimes eg)$.)
- (iii) $(f \otimes ge)\Delta^{(-1)} = (f \otimes g)\Delta^{(-1)}e$ (and $= (fe \otimes g)\Delta^{(-1)}$ and $= (fe \otimes ge)\Delta^{(-1)}$.)
- (iv) $\overline{\Delta(f \otimes g)\Delta^{(-1)}} = \Delta(1 \otimes gf^{(-1)})\Delta^{(-1)}$.
- (v) If $\Delta(h \otimes g)\Delta^{(-1)} = \overline{\Delta(h \otimes g)\Delta^{(-1)}}$ then $(\Delta(h \otimes g)\Delta^{(-1)})h = \Delta(h \otimes g)\Delta^{(-1)}$.
- (vi) $\Delta(f \otimes 1) = \Delta(g \otimes 1) \implies f = g$.
- (vii) $(f \otimes 1) = (g \otimes 1) \implies f = g$.

Proof.

(i) This is shown by proving both sides equal $\Delta(e \otimes 1)\Delta^{(-1)}\Delta(e \otimes 1)\Delta^{(-1)}$.

$$\begin{aligned}
& \Delta(e \otimes 1)\Delta^{(-1)}\Delta(e \otimes 1)\Delta^{(-1)} = \Delta(e \otimes 1)\Delta^{(-1)}\Delta(1 \otimes e)\Delta^{(-1)} && \text{cocommutativity} \\
& = \Delta(e\Delta \otimes 1)(1 \otimes \Delta^{(-1)}e)\Delta^{(-1)} && \text{Frobenius} \\
& = \Delta(\Delta \otimes 1)(e \otimes e \otimes 1)(1 \otimes \Delta^{(-1)}e)\Delta^{(-1)} && \Delta \text{ natural} \\
& = \Delta(\Delta \otimes 1)(e \otimes e \otimes 1)(1 \otimes e \otimes e)(1 \otimes \Delta^{(-1)})\Delta^{(-1)} && \Delta^{(-1)} \text{ natural} \\
& = \Delta(\Delta \otimes 1)(e \otimes e \otimes e)(1 \otimes \Delta^{(-1)})\Delta^{(-1)} && e \text{ idempotent} \\
& = \Delta(\Delta \otimes 1)(e \otimes \Delta^{(-1)}e)\Delta^{(-1)} && \Delta^{(-1)} \text{ natural} \\
& = \Delta(\Delta \otimes 1)(1 \otimes \Delta^{(-1)})\Delta^{(-1)}e && \Delta^{(-1)} \text{ natural} \\
& = \Delta\Delta^{(-1)}\Delta\Delta^{(-1)}e && \text{Frobenius} \\
& = e && \Delta \text{ total.}
\end{aligned}$$

At the same time,

$$\begin{aligned}
\Delta(e \otimes 1)\Delta^{(-1)}\Delta(e \otimes 1)\Delta^{(-1)} &= \Delta(e\Delta \otimes 1)(e \otimes \Delta^{(-1)}1)\Delta^{(-1)} && \text{Frobenius} \\
&= \Delta(\Delta \otimes 1)(e \otimes e \otimes 1)(e \otimes \Delta^{(-1)})\Delta^{(-1)} && \Delta_{\text{natural}} \\
&= \Delta(\Delta \otimes 1)(e \otimes e \otimes 1)(e \otimes 1 \otimes 1)(1 \otimes \Delta^{(-1)})\Delta^{(-1)} && \Delta^{(-1)} \text{ natural} \\
&= \Delta(\Delta \otimes 1)(e \otimes e \otimes 1)(1 \otimes \Delta^{(-1)})\Delta^{(-1)} && e \text{ idempotent} \\
&= \Delta(e\Delta \otimes 1)(1 \otimes \Delta^{(-1)})\Delta^{(-1)} && \Delta_{\text{natural}} \\
&= \Delta(e \otimes 1)\Delta^{(-1)}\Delta\Delta^{(-1)} && \text{Frobenius} \\
&= \Delta(e \otimes 1)\Delta^{(-1)} && \Delta \text{ total}
\end{aligned}$$

which gives $e = \Delta(e \otimes 1)\Delta^{(-1)}$.

(ii) This equality starts by using the previous equality:

$$\begin{aligned}
e\Delta(f \otimes g) &= \Delta(e \otimes 1)\Delta^{(-1)}\Delta(f \otimes g) && \text{by part (i)} \\
&= \Delta(e \otimes 1)\overline{\Delta^{(-1)}}(f \otimes g) \\
&= \Delta\overline{\Delta^{(-1)}}(e \otimes 1)(f \otimes g) && [\mathbf{R.2}] \text{ as } e \otimes 1 \text{ is a restriction idempotent} \\
&= \Delta(ef \otimes g) && (ff^{(-1)} = f).
\end{aligned}$$

The second and third equalities follow by cocommutativity, naturality of Δ and e being a restriction idempotent.

(iii) As in ((ii) on the preceding page), details are only given for the first equality.

$$\begin{aligned}
(f \otimes g)\Delta^{(-1)}e &= (f \otimes g)\Delta^{(-1)}\Delta(1 \otimes e)\Delta^{(-1)} && \text{part (i)} \\
&= (f \otimes g)\overline{\Delta^{(-1)}}(1 \otimes e)\Delta^{(-1)} \\
&= (f \otimes g)(1 \otimes e)\overline{\Delta^{(-1)}}\Delta^{(-1)} && [\mathbf{R.2}] \\
&= (f \otimes ge)\Delta^{(-1)} && [\mathbf{R.1}]
\end{aligned}$$

The other equalities follow from co-commutativity, naturality of Δ and e being a restriction idempotent.

(iv) Here, we start by using the fact all maps have a partial inverse:

$$\begin{aligned}
& \overline{\Delta(f \otimes g) \Delta^{(-1)}} \\
&= \Delta(f \otimes g) \Delta^{(-1)} \Delta(f^{(-1)} \otimes g^{(-1)}) \Delta^{(-1)} \\
&= \Delta(g \otimes f) \Delta^{(-1)} \Delta(g^{(-1)} \otimes f^{(-1)}) \Delta^{(-1)} && \text{co-commutative} \\
&= \Delta(g \Delta \otimes f) (g^{(-1)} \otimes \Delta^{(-1)} f^{(-1)}) \Delta^{(-1)} && \text{Frobenius} \\
&= \Delta(\Delta \otimes 1) (g \otimes g \otimes f) (g^{(-1)} \otimes \Delta^{(-1)} f^{(-1)}) \Delta^{(-1)} && \Delta \text{ natural} \\
&= \Delta(\Delta \otimes 1) (g \otimes g \otimes f) (g^{(-1)} \otimes f^{(-1)} \otimes f^{(-1)}) (1 \otimes \Delta^{(-1)}) \Delta^{(-1)} && \Delta^{(-1)} \text{ natural} \\
&= \Delta(\Delta \otimes 1) (\bar{g} \otimes g f^{(-1)} \otimes \bar{f}) (1 \otimes \Delta^{(-1)}) \Delta^{(-1)} && \text{combine maps} \\
&= \Delta(\Delta \otimes 1) (\bar{g} \otimes \bar{g} g f^{(-1)} \bar{f} \otimes \bar{f}) (1 \otimes \Delta^{(-1)}) \Delta^{(-1)} && \text{restriction axioms} \\
&= \Delta(\bar{g} \Delta \otimes 1) (1 \otimes g f^{(-1)} \bar{f} \otimes \bar{f}) (1 \otimes \Delta^{(-1)}) \Delta^{(-1)} && \Delta \text{ natural} \\
&= \Delta(\bar{g} \Delta \otimes 1) (1 \otimes g f^{(-1)} \otimes 1) (1 \otimes \Delta^{(-1)} \bar{f}) \Delta^{(-1)} && \Delta^{(-1)} \text{ natural} \\
&= \Delta(\Delta \otimes 1) (1 \otimes \bar{g} g f^{(-1)} \otimes 1) (1 \otimes \Delta^{(-1)} \bar{f}) \Delta^{(-1)} && \text{This lemma((ii))} \\
&= \Delta(\Delta \otimes 1) (1 \otimes \bar{g} g f^{(-1)} \bar{f} \otimes 1) (1 \otimes \Delta^{(-1)}) \Delta^{(-1)} && \text{This lemma((iii))} \\
&= \Delta(\Delta \otimes 1) (1 \otimes g f^{(-1)} \otimes 1) (1 \otimes \Delta^{(-1)}) \Delta^{(-1)} && \text{restriction axioms} \\
&= \Delta c_{A,A} (\Delta \otimes 1) (1 \otimes g f^{(-1)} \otimes 1) (1 \otimes \Delta^{(-1)}) \Delta^{(-1)} && \text{co-commutative} \\
&= \Delta(1 \otimes \Delta) c_{A,A \otimes A} (1 \otimes g f^{(-1)} \otimes 1) (1 \otimes \Delta^{(-1)}) \Delta^{(-1)} && c_{\otimes} \text{ natural} \\
&= \Delta(1 \otimes \Delta) (1 \otimes 1 \otimes g f^{(-1)}) c_{A,A \otimes A} (1 \otimes \Delta^{(-1)}) \Delta^{(-1)} && c_{\otimes} \text{ natural} \\
&= \Delta(1 \otimes \Delta) (1 \otimes 1 \otimes g f^{(-1)}) (\Delta^{(-1)} \otimes 1) c_{A,A} \Delta^{(-1)} && c_{\otimes} \text{ natural} \\
&= \Delta(1 \otimes \Delta) (1 \otimes 1 \otimes g f^{(-1)}) (\Delta^{(-1)} \otimes 1) \Delta^{(-1)} && c_{\otimes} \text{ co-commutative} \\
&= \Delta \Delta^{(-1)} \Delta (1 \otimes g f^{(-1)}) \Delta^{(-1)} && \text{Frobenius} \\
&= \Delta(1 \otimes g f^{(-1)}) \Delta^{(-1)} && \Delta \text{ total}
\end{aligned}$$

Note the pattern in the last few lines of using the co-commutativity of Δ ,

the naturality of the commutativity isomorphism and finishing with the co-commutativity of $\Delta^{(-1)}$. In future proofs, these steps will be combined to a single line and referred to as commutativity.

(v) Beginning with the assumption,

$$\begin{aligned}
(\Delta(h \otimes g)\Delta^{(-1)})h &= \overline{\Delta(h \otimes g)\Delta^{(-1)}}h \\
&= \Delta(1 \otimes gh^{(-1)})\Delta^{(-1)}h && \text{This lemma((iv))} \\
&= \Delta(1 \otimes gh^{(-1)})\Delta^{(-1)}\Delta(h \otimes h)\Delta^{(-1)} && \Delta \text{ total and natural} \\
&= \Delta(1 \otimes gh^{(-1)})(\Delta \otimes 1)(1 \otimes \Delta^{(-1)})(h \otimes h)\Delta^{(-1)} && \text{Frobenius} \\
&= \Delta(\Delta \otimes 1)(1 \otimes 1 \otimes gh^{(-1)})(1 \otimes \Delta^{(-1)})(h \otimes h)\Delta^{(-1)} && \Delta \text{ natural} \\
&= \Delta(\Delta \otimes 1)(1 \otimes 1 \otimes gh^{(-1)})(h \otimes h \otimes h)(1 \otimes \Delta^{(-1)})\Delta^{(-1)} && \Delta^{(-1)} \text{ natural} \\
&= \Delta(\Delta \otimes 1)(h \otimes h \otimes gh^{(-1)}h)(1 \otimes \Delta^{(-1)})\Delta^{(-1)} && \text{combine terms} \\
&= \Delta(h \otimes g\overline{h^{(-1)}})(\Delta \otimes 1)(1 \otimes \Delta^{(-1)})\Delta^{(-1)} && \Delta \text{ natural} \\
&= \Delta(h \otimes g\overline{h^{(-1)}})\Delta^{(-1)}\Delta\Delta^{(-1)} && \text{Frobenius} \\
&= \Delta(h \otimes g\overline{h^{(-1)}})\Delta^{(-1)} && \Delta \text{ total} \\
&= \Delta(h \otimes g)\Delta^{(-1)}\overline{h^{(-1)}} && \text{part ((ii))} \\
&= \Delta(\overline{hh^{(-1)}} \otimes g)\Delta^{(-1)} && \text{part ((ii))} \\
&= \Delta(h \otimes g)\Delta^{(-1)} && \text{property of inverse.}
\end{aligned}$$

(vi) As Δ is total and natural, we start with:

$$\begin{aligned}
f &= \Delta(f \otimes f)\Delta^{(-1)} \\
&= \Delta(f \otimes 1)(1 \otimes f)\Delta^{(-1)} \\
&= \Delta(g \otimes 1)(1 \otimes f)\Delta^{(-1)} && \text{assumption} \\
&= \Delta(1 \otimes f)(g \otimes 1)\Delta^{(-1)} && \text{Identities commute} \\
&= \Delta(1 \otimes g)(g \otimes 1)\Delta^{(-1)} && \text{assumption, co-commutative} \\
&= \Delta(g \otimes g)\Delta^{(-1)} \\
&= g\Delta\Delta^{(-1)} && \Delta \text{ natural} \\
&= g && \Delta \text{ total.}
\end{aligned}$$

(vii) Immediate from part (vi) on page 38.

□

Proposition 3.1.9. *A discrete inverse category has meets, where $f \cap g = \Delta(f \otimes g)\Delta^{(-1)}$.*

Proof. $f \cap g \leq f$:

$$\begin{aligned}
f \cap g &= \Delta(f \otimes g)\Delta^{(-1)} && \text{Definition of } \cap \\
&= \Delta(\overline{ff^{(-1)}} \otimes g)\Delta^{(-1)} && \text{property of inverse} \\
&= \Delta(f \otimes g\overline{f^{(-1)}})\Delta^{(-1)} && \text{by lemma 3.1.8((iii))} \\
&= \Delta(f \otimes gf^{(-1)}f)\Delta^{(-1)} && \text{definition of inverse} \\
&= \Delta(1 \otimes gf^{(-1)})\Delta^{(-1)}f && \Delta^{(-1)} \text{ natural} \\
&= \overline{f \cap g}f && \text{by lemma 3.1.8((iv))}
\end{aligned}$$

$$f \cap f = f:$$

$$\begin{aligned} f \cap f &= \Delta(f \otimes f) \Delta^{(-1)} \\ &= f \Delta \Delta^{(-1)} && \Delta \text{ natural} \\ &= f && \Delta \text{ total.} \end{aligned}$$

$$h(f \cap g) = hf \cap hg:$$

$$\begin{aligned} h(f \cap g) &= h \Delta(f \otimes g) \Delta^{(-1)} && \text{Definition of } \cap \\ &= \Delta(h \otimes h)(f \otimes g) \Delta^{(-1)} && \Delta \text{ natural} \\ &= \Delta(hf \otimes hg) \Delta^{(-1)} && \text{compose maps} \\ &= hf \cap hg && \text{Definition of } \cap. \end{aligned}$$

□

3.1.4 The inverse subcategory of a discrete restriction category

Given a discrete restriction category, one can pick out the maps which are partial isomorphisms. Using results from the previous sub-section and from sub-section [sub-section 2.2.7 on page 29](#), this section will show that these maps form a restriction subcategory and in fact, form a discrete inverse category.

Lemma 3.1.10. *Given \mathbb{X} is a discrete restriction category, the invertible maps of \mathbb{X} , together with the objects of \mathbb{X} form a sub restriction category which is a discrete inverse category, denoted by $Inv(\mathbb{X})$.*

Proof. As shown in Lemma [2.2.13 on page 20](#), partial isomorphisms are closed under composition. The identity maps are in $Inv(\mathbb{X})$. Trivially, restrictions of partial isomorphisms are also partial isomorphisms.

The product on the discrete restriction category \mathbb{X} becomes the tensor product of the restriction category $Inv(\mathbb{X})$. Table [table 3.1 on the following page](#) shows how each of the

elements of the tensor are defined. Note that the last definition makes explicit use of the fact we are in a discrete restriction category and hence the Δ of \mathbb{X} possesses a partial inverse.

\mathbb{X}	$Inv(\mathbb{X})$	Inverse map
$A \times B$	$A \otimes B$	
\top	1	
$\pi_1: \top \times A \rightarrow A$	$u_{\otimes}^l: 1 \otimes A \rightarrow A$	$\langle !, 1 \rangle$
$\pi_0: A \times \top \rightarrow A$	$u_{\otimes}^r: A \otimes 1 \rightarrow A$	$\langle 1, ! \rangle$
$\langle \pi_0 \pi_0, \langle \pi_0 \pi_1, \pi_1 \rangle \rangle: (A \times B) \times C \rightarrow A \times (B \times C)$	$a_{\otimes}: (A \otimes B) \otimes C \rightarrow A \otimes (B \otimes C)$	$\langle \langle \pi_0, \pi_1 \pi_0 \rangle, \pi_1 \pi_1 \rangle$
$\langle \pi_1, \pi_0 \rangle: A \times B \rightarrow B \times A$	$c_{\otimes}: A \otimes B \rightarrow B \otimes A$	$\langle \pi_1, \pi_0 \rangle$
$\Delta_{\mathbb{X}}: A \rightarrow A \times A$	$\Delta: A \rightarrow A \otimes A$	$\Delta_{\mathbb{X}}^{(-1)}$

Table 3.1: Structural maps for the tensor in $Inv(\mathbb{X})$

The monoid coherence diagrams and Δ being total follow directly from the characteristics of the product in \mathbb{X} . It remains to show co-commutativity, co-associativity and the Frobenius condition.

Co-commutativity requires $\Delta c_{\otimes} = c_{\otimes}$. From the definitions, this means we need

$$\Delta_{\mathbb{X}} \langle \pi_1, \pi_0 \rangle = \Delta_{\mathbb{X}}.$$

Once again, this follows immediately from the definition of restriction product.

Co-associativity requires $\Delta(1 \otimes \Delta) = \Delta(\Delta \otimes 1)a_{\otimes}$. Expressing this in \mathbb{X} , we require

$$\Delta_{\mathbb{X}}(1 \times \Delta_{\mathbb{X}}) = \Delta_{\mathbb{X}}(\Delta_{\mathbb{X}} \times 1)(\langle \pi_0 \pi_0, \langle \pi_0 \pi_1, \pi_1 \rangle \rangle).$$

Again each is equal based on the properties of the restriction product.

The Frobenius requirement is two-fold:

$$\Delta^{(-1)} \Delta = (\Delta \otimes 1)a_{\otimes}(1 \otimes \Delta^{(-1)}) \quad (3.1)$$

$$\Delta^{(-1)} \Delta = (1 \otimes \Delta)a_{\otimes}^{(-1)}(\Delta^{(-1)} \otimes 1), \quad (3.2)$$

but in \mathbb{X} , this becomes:

$$\Delta_{\mathbb{X}}^{(-1)} \Delta_{\mathbb{X}} = (\Delta_{\mathbb{X}} \times 1) \langle \pi_0 \pi_0, \langle \pi_0 \pi_1, \pi_1 \rangle \rangle (1 \times \Delta_{\mathbb{X}}^{(-1)}) \quad (3.3)$$

$$\Delta_{\mathbb{X}}^{(-1)} \Delta_{\mathbb{X}} = (1 \times \Delta_{\mathbb{X}}) \langle \langle \pi_0, \pi_1 \pi_0 \rangle, \pi_1 \pi_1 \rangle (\Delta_{\mathbb{X}}^{(-1)} \times 1). \quad (3.4)$$

We will detail the proof of equation [equation \(3.3\) on the previous page](#). Equation [equation \(3.4\) on the preceding page](#) is proved similarly.

To show the equation, note first that $\Delta(1 \times !)$ (and $\Delta(! \times 1)$) is the identity and secondly that maps to a product of objects may be split into a product map — e.g. if $f : A \rightarrow B \times B$, then $f = \langle f(1 \times !), f(! \times 1) \rangle$.

Using this we see that the left hand side of equation [equation \(3.3\) on the previous page](#) computes as follows:

$$\begin{aligned}\Delta_{\mathbb{X}}^{(-1)}\Delta_{\mathbb{X}} &= \langle \Delta_{\mathbb{X}}^{(-1)}\Delta_{\mathbb{X}}(1 \times !), \Delta_{\mathbb{X}}^{(-1)}\Delta_{\mathbb{X}}(! \times 1) \rangle \\ &= \langle \Delta_{\mathbb{X}}^{(-1)}, \Delta_{\mathbb{X}}^{(-1)} \rangle\end{aligned}$$

Similarly, removing the associativity maps, the right hand side of the same equation becomes:

$$\begin{aligned}(\Delta_{\mathbb{X}} \times 1)(1 \times \Delta_{\mathbb{X}}^{(-1)}) &= \langle (\Delta_{\mathbb{X}} \times 1)(1 \times \Delta_{\mathbb{X}}^{(-1)})(1 \times !), (\Delta_{\mathbb{X}} \times 1)(1 \times \Delta_{\mathbb{X}}^{(-1)})(! \times 1) \rangle \\ &= \langle (\Delta_{\mathbb{X}} \times 1)(1 \times \Delta_{\mathbb{X}}^{(-1)})(1 \times !), \Delta_{\mathbb{X}}^{(-1)} \rangle \\ &= \langle (\Delta_{\mathbb{X}} \times 1)(1 \times \Delta_{\mathbb{X}}^{(-1)})(1 \times \Delta_{\mathbb{X}})(1 \times ! \times !), \Delta_{\mathbb{X}}^{(-1)} \rangle \\ &= \langle (\Delta_{\mathbb{X}} \times 1)(1 \times \overline{\Delta_{\mathbb{X}}^{(-1)}})(1 \times ! \times !), \Delta_{\mathbb{X}}^{(-1)} \rangle \\ &= \langle (\Delta_{\mathbb{X}} \times 1)\overline{1 \times \Delta_{\mathbb{X}}^{(-1)}}(1 \times ! \times !), \Delta_{\mathbb{X}}^{(-1)} \rangle \\ &= \langle \overline{(\Delta_{\mathbb{X}} \times 1)(1 \times \Delta_{\mathbb{X}}^{(-1)})}(\Delta_{\mathbb{X}} \times 1)(1 \times ! \times !), \Delta_{\mathbb{X}}^{(-1)} \rangle \\ &= \langle \overline{(\Delta_{\mathbb{X}} \times 1)(1 \times \Delta_{\mathbb{X}}^{(-1)})}(1 \times !), \Delta_{\mathbb{X}}^{(-1)} \rangle \\ &= \langle \overline{(\Delta_{\mathbb{X}} \times 1)(1 \times \Delta_{\mathbb{X}}^{(-1)})}(! \times 1)(1 \times !), \Delta_{\mathbb{X}}^{(-1)} \rangle \\ &= \langle \overline{\Delta_{\mathbb{X}}^{(-1)}}(1 \times !), \Delta_{\mathbb{X}}^{(-1)} \rangle \\ &= \langle \Delta_{\mathbb{X}}^{(-1)}\Delta_{\mathbb{X}}(1 \times !), \Delta_{\mathbb{X}}^{(-1)} \rangle \\ &= \langle \Delta_{\mathbb{X}}^{(-1)}, \Delta_{\mathbb{X}}^{(-1)} \rangle\end{aligned}$$

and therefore we see that the first equation for the Frobenius condition is satisfied. Thus, $Inv(\mathbb{X})$ is a discrete inverse category. □

3.2 Completing a discrete inverse category

The purpose of this section is to prove that the category of discrete inverse categories is equivalent to the the category of discrete restriction categories. In order to prove this, we show how to construct a discrete restriction category, $\widetilde{\mathbb{X}}$, from a discrete inverse category, \mathbb{X} .

3.2.1 The restriction category $\widetilde{\mathbb{X}}$

Definition 3.2.1. When \mathbb{X} is an inverse category, define $\widetilde{\mathbb{X}}$ as:

Objects: objects as in \mathbb{X}

Maps: equivalence classes of maps (the equivalence class is defined below in Definition 3.2.2 on the following page) with the following structure in \mathbb{X} :

$$\frac{A \xrightarrow{(f,C)} B \text{ in } \widetilde{\mathbb{X}}}{A \xrightarrow{f} B \otimes C \text{ in } \mathbb{X}}$$

Identity: by

$$\frac{A \xrightarrow{(u_{\otimes}^r(-1),1)} A}{A \xrightarrow{u_{\otimes}^r(-1)} A \otimes 1}$$

Composition: given by

$$\frac{\frac{A \xrightarrow{(f,B')} B \xrightarrow{(g,C')} C}{A \xrightarrow{f(g \otimes 1)a_{\otimes}} C \otimes (C' \otimes B')}}{A \xrightarrow{(f(g \otimes 1)a_{\otimes}, C' \otimes B')} C}$$

When considering an $\widetilde{\mathbb{X}}$ map $(f, C) : A \rightarrow B$ in \mathbb{X} , we occasionally use the notation $f : A \rightarrow B|_C (\equiv f : A \rightarrow B \otimes C)$.

Equivalence classes of maps in \mathbb{X}

Definition 3.2.2. In a discrete inverse category \mathbb{X} as defined above, the map f is equivalent to f' in \mathbb{X} when $\bar{f} = \bar{f}'$ in \mathbb{X} and the below diagram commutes for some map h :

$$\begin{array}{ccccc}
 & & B \otimes C & \xrightarrow{(\Delta \otimes 1) a_{\otimes}} & B \otimes (B \otimes C) \\
 & \nearrow f & & & \downarrow 1 \otimes h \\
 A & & & & B \otimes (B \otimes C') \\
 & \searrow f' & & \xleftarrow{a_{\otimes}^{(-1)} (\Delta^{(-1)} \otimes 1)} & \\
 & & B \otimes C' & &
 \end{array}$$

Notation 3.2.3. When f is equivalent to g via the mediating map h , this is written as

$$f \stackrel{h}{\simeq} g.$$

Lemma 3.2.4. Definition 3.2.2 gives a symmetric, reflexive equivalence class of maps in \mathbb{X} .

Proof.

Reflexivity: Choose h as the identity map.

Symmetry: Suppose $f \stackrel{h}{\simeq} g$. Then, $\bar{f} = \bar{g}$ and $fk = g$ where

$$k = (\Delta \otimes 1) a_{\otimes} (1 \otimes h) a_{\otimes}^{(-1)} (\Delta^{(-1)} \otimes 1).$$

Applying $k^{(-1)}$, which is

$$(\Delta \otimes 1) a_{\otimes} (1 \otimes h^{(-1)}) a_{\otimes}^{(-1)} (\Delta^{(-1)} \otimes 1),$$

we have

$$gk^{(-1)} = fkk^{(-1)} = f\bar{k} = \bar{f}k f = \bar{g}f = \bar{f}f = f.$$

Thus, $g \stackrel{h^{(-1)}}{\simeq} f$.

Transitivity: Suppose $f \xrightarrow{h} f'$ and $f' \xrightarrow{k} f''$. Then, consider the compositions of the mediating portions of the equivalences:

$$\ell = ((\Delta \otimes 1)a_{\otimes}(1 \otimes h)a_{\otimes}^{(-1)}(\Delta^{(-1)} \otimes 1))((\Delta \otimes 1)a_{\otimes}(1 \otimes k)a_{\otimes}^{(-1)}(\Delta^{(-1)} \otimes 1)).$$

By pasting the diagrams which give the above equivalences, we see that $f\ell = f''$. However, it is not in the form of a mediating map as presented.

The claim is that ℓ is the actual mediating map for f and f'' . That is, that we have $f(\Delta \otimes 1)a_{\otimes}(1 \otimes \ell)a_{\otimes}^{(-1)}(\Delta^{(-1)} \otimes 1) = f''$. In the interest of some brevity, this is shown below with the associativity maps elided from the equations.

We need to show that $(\Delta \otimes 1)(1 \otimes \ell)(\Delta^{(-1)} \otimes 1) = \ell$.

$$\begin{aligned} & (\Delta \otimes 1)(1 \otimes \ell)(\Delta^{(-1)} \otimes 1) \\ &= (\Delta \otimes 1)(1 \otimes \Delta \otimes 1)(1 \otimes 1 \otimes h)(1 \otimes \Delta^{(-1)} \otimes 1) \\ & \quad (1 \otimes \Delta \otimes 1)(1 \otimes 1 \otimes k)(1 \otimes \Delta^{(-1)} \otimes 1)(\Delta^{(-1)} \otimes 1) \\ &= (\Delta \otimes 1)(\Delta \otimes 1 \otimes 1)(1 \otimes 1 \otimes h)(1 \otimes \Delta^{(-1)} \otimes 1) \\ & \quad (1 \otimes \Delta \otimes 1)(1 \otimes 1 \otimes k)(\Delta^{(-1)} \otimes 1 \otimes 1)(\Delta^{(-1)} \otimes 1) \quad \text{co-associativity} \\ &= (\Delta \otimes 1)(1 \otimes h)(\Delta \otimes 1 \otimes 1)(1 \otimes \Delta^{(-1)} \otimes 1) \\ & \quad (1 \otimes \Delta \otimes 1)(\Delta^{(-1)} \otimes 1 \otimes 1)(1 \otimes k)(\Delta^{(-1)} \otimes 1) \quad \text{Naturality} \\ &= (\Delta \otimes 1)(1 \otimes h)(\Delta^{(-1)} \otimes 1)(\Delta \otimes 1) \\ & \quad (\Delta^{(-1)} \otimes 1)(\Delta \otimes 1)(1 \otimes k)(\Delta^{(-1)} \otimes 1) \quad \text{Frobenius} \\ &= (\Delta \otimes 1)(1 \otimes h)(\Delta^{(-1)} \otimes 1)(\Delta \otimes 1)(1 \otimes k)(\Delta^{(-1)} \otimes 1) \quad \Delta \text{ Total} \\ &= \ell \end{aligned}$$

and therefore $f \xrightarrow{\ell} f''$. □

Corollary 3.2.5. *If $\bar{f} = \bar{g}$ in \mathbb{X} , a discrete inverse category, and the diagram*

$$\begin{array}{ccc} & & B \otimes C \\ & \nearrow f & \downarrow 1 \otimes h \\ A & & B \otimes C' \\ & \searrow g & \end{array}$$

commutes for some h , then there is a h' such that $f \stackrel{h'}{\simeq} g$.

Proof. Consider

$$\begin{aligned} (\Delta \otimes 1) a_{\otimes} (1 \otimes (1 \otimes h)) a_{\otimes}^{(-1)} (\Delta^{(-1)} \otimes 1) & \\ = (\Delta \otimes 1) ((1 \otimes 1) \otimes h) a_{\otimes} a_{\otimes}^{(-1)} (\Delta^{(-1)} \otimes 1) & \text{Naturality} \\ = (\Delta \otimes 1) ((1 \otimes 1) \otimes h) (\Delta^{(-1)} \otimes 1) & \text{Isomorphism Inverse} \\ = (\Delta(1 \otimes 1) \Delta^{(-1)}) \otimes h & \text{Naturality of } \otimes \\ = (1 \otimes h) & \Delta \Delta^{(-1)} = 1 \end{aligned}$$

and therefore we can set $h' = 1 \otimes h$. □

Lemma 3.2.6. $\widetilde{\mathbb{X}}$ as defined above is a category.

Proof. The maps are well defined, as shown in lemma 3.2.4 on page 47. The existence of the identity map is due to the tensor \otimes being defined on \mathbb{X} , an inverse category, hence $u_{\otimes}^{r(-1)}$ is defined.

It remains to show the composition is associative and that $(u_{\otimes}^{r(-1)}, 1)$ acts as an identity in $\widetilde{\mathbb{X}}$.

Associativity: Consider

$$A \xrightarrow{(f, B')} B \xrightarrow{(g, C')} C \xrightarrow{(h, D')} D.$$

To show the associativity of this in $\widetilde{\mathbb{X}}$, we need to show in \mathbb{X} that

$$\overline{(f(g \otimes 1) a_{\otimes})(h \otimes 1) a_{\otimes}} = \overline{f(((g(h \otimes 1) a_{\otimes}) \otimes 1) a_{\otimes})}$$

From this, we can conclude

$$(f(g \otimes 1)a_{\otimes})(h \otimes 1)a_{\otimes} \stackrel{1 \otimes a_{\otimes}^{(-1)}}{\simeq} f(((g(h \otimes 1)a_{\otimes}) \otimes 1)a_{\otimes})$$

which gives us that composition in $\widetilde{\mathbb{X}}$ is associative.

Identity: This requires:

$$(f, C)(u_{\otimes}^{r(-1)}, 1) = (f, C) = (u_{\otimes}^{r(-1)}, 1)(f, C)$$

for all maps $A \xrightarrow{(f, C)} B$ in $\widetilde{\mathbb{X}}$.

First, we see $\overline{f(u_{\otimes}^{r(-1)} \otimes 1)a_{\otimes}} = \overline{f}$ by Lemma 2.2.3 on page 14. Then, calculating in \mathbb{X} , we have a mediating map of $1 \otimes u_{\otimes}^l$ as shown below.

$$\begin{array}{ccccccc}
 A & \xrightarrow{f} & B \otimes C & \xrightarrow{u_{\otimes}^{r(-1)} \otimes 1} & (B \otimes 1) \otimes C & \xrightarrow{a_{\otimes}} & B \otimes (1 \otimes C) \\
 & & & \searrow & \searrow & & \downarrow 1 \otimes u_{\otimes}^l \\
 & & & & & & B \otimes C \\
 & \searrow f & & & & & \\
 & & & & & &
 \end{array}$$

$\xrightarrow{1 \otimes u_{\otimes}^l} B \otimes C$

Next, $\overline{u_{\otimes}^{r(-1)}(f \otimes 1)a_{\otimes}} = \overline{f}$ by the naturality of $u_{\otimes}^{r(-1)}$ and Lemma 2.2.3 on page 14. The diagram below

$$\begin{array}{ccccccc}
 A & \xrightarrow{u_{\otimes}^{r(-1)}} & A \otimes 1 & \xrightarrow{f \otimes 1} & (B \otimes C) \otimes 1 & \xrightarrow{a_{\otimes}} & B \otimes (C \otimes 1) \\
 & \searrow f & & \nearrow u_{\otimes}^{r(-1)} & \nearrow 1 \otimes u_{\otimes}^{r(-1)} & & \downarrow 1 \otimes u_{\otimes}^r \\
 & & & B \otimes C & & & B \otimes C \\
 & \searrow f & & & & & \\
 & & & & & &
 \end{array}$$

$\xrightarrow{1 \otimes u_{\otimes}^r} B \otimes C$

shows our mediating map is $1 \otimes u_{\otimes}^r$. □

Defining the restriction on $\widetilde{\mathbb{X}}$

Define the restriction in $\widetilde{\mathbb{X}}$ as follows:

$$\frac{\frac{A \xrightarrow{(f,C)} B}{A \xrightarrow{(f,C)} A}}{A \xrightarrow{\bar{f}u_{\otimes}^{r(-1)}} A \otimes 1 \text{ in } \mathbb{X}}$$

Lemma 3.2.7. *The category $\widetilde{\mathbb{X}}$ with restriction defined as above is a restriction category.*

Proof. Given the above definition, the four restriction axioms must now be checked. (Diagrams are in \mathbb{X}).

[R.1] ($\bar{f}f = f$) Calculating the restriction of the left hand side in \mathbb{X} , we have:

$$\begin{aligned} \overline{\bar{f}u_{\otimes}^{r(-1)}(f \otimes 1)a_{\otimes}} &= \overline{\bar{f}u_{\otimes}^{r(-1)}(f \otimes 1)} && a_{\otimes} \text{ iso, Lemma 2.2.3} \\ &= \overline{\bar{f}fu_{\otimes}^{r(-1)}} && u_{\otimes}^{r(-1)} \text{ natural} \\ &= \overline{fu_{\otimes}^{r(-1)}} && [\text{R.1}] \text{ in } \mathbb{X} \\ &= \bar{f} && u_{\otimes}^{r(-1)} \text{ iso, Lemma 2.2.3.} \end{aligned}$$

Then, the following diagram

$$\begin{array}{ccccccc} A & \xrightarrow{\bar{f}u_{\otimes}^{r(-1)}} & A \otimes 1 & \xrightarrow{f \otimes 1} & (A \otimes B) \otimes 1 & \xrightarrow{a_{\otimes}} & A \otimes (B \otimes 1) \\ & \searrow \bar{f}f & & \nearrow u_{\otimes}^{r(-1)} & \searrow u_{\otimes}^r & & \downarrow 1 \otimes u_{\otimes}^r \\ & & & A \otimes B & & & A \otimes B \\ & \searrow f & & & & & \downarrow \\ & & & & & & A \otimes B \end{array}$$

shows $\bar{f}u_{\otimes}^{r(-1)}(f \otimes 1)a_{\otimes} \xrightarrow{1 \otimes u_{\otimes}^r} f$ in \mathbb{X} and therefore $\bar{f}f = f$ in $\widetilde{\mathbb{X}}$.

[R.2] ($\bar{g}\bar{f} = \bar{f}\bar{g}$) The restriction of the left hand side equals the restriction of the right hand side as seen below:

$$\begin{aligned} \overline{\bar{f}u_{\otimes}^{r(-1)}((\bar{g}u_{\otimes}^{r(-1)}) \otimes 1)a_{\otimes}} &= \overline{\bar{f}(\bar{g}u_{\otimes}^{r(-1)})u_{\otimes}^{r(-1)}a_{\otimes}} && u_{\otimes}^{r(-1)} \text{ natural} \\ &= \overline{\bar{g}\bar{f}u_{\otimes}^{r(-1)}u_{\otimes}^{r(-1)}a_{\otimes}} && [\text{R.2}] \text{ in } \mathbb{X} \\ &= \overline{\bar{g}u_{\otimes}^{r(-1)}((\bar{f}u_{\otimes}^{r(-1)}) \otimes 1)a_{\otimes}} && u_{\otimes}^{r(-1)} \text{ natural.} \end{aligned}$$

The below diagram commutes by the naturality of u_{\otimes}^r and the tensor coherence,

$$\begin{array}{ccccc}
A & \xrightarrow{\bar{g}u_{\otimes}^{r(-1)}} & A \otimes 1 & \xrightarrow{(\bar{f}u_{\otimes}^{r(-1)}) \otimes 1} & (A \otimes 1) \otimes 1 \xrightarrow{a_{\otimes}} A \otimes (1 \otimes 1) \\
\downarrow \bar{f}u_{\otimes}^{r(-1)} & \searrow \bar{g}\bar{f} & & \nearrow u_{\otimes}^r u_{\otimes}^r & \downarrow 1 \otimes id \\
A \otimes 1 & & A & \nwarrow u_{\otimes}^{r(-1)} u_{\otimes}^{r(-1)} & \\
\downarrow (\bar{g}u_{\otimes}^{r(-1)}) \otimes 1 & \nearrow u_{\otimes}^r u_{\otimes}^r & & & \\
(A \otimes 1) \otimes 1 & \xrightarrow{a_{\otimes}} & A \otimes (1 \otimes 1) & &
\end{array}$$

which allows us to conclude $\bar{f}\bar{g} = \bar{g}\bar{f}$ in $\widetilde{\mathbb{X}}$.

R.3 ($\overline{\bar{f}g} = \bar{f}\bar{g}$). As above, the first step is to show that the restrictions of each side are the same. Computing the restriction of the left hand side in \mathbb{X} :

$$\begin{aligned}
\overline{(\bar{f}u_{\otimes}^{r(-1)})(g \otimes 1)a_{\otimes}u_{\otimes}^{r(-1)}} &= \overline{(\bar{f}u_{\otimes}^{r(-1)})(g \otimes 1)a_{\otimes}} && u_{\otimes}^{r(-1)} \text{ iso, Lemma 2.2.3} \\
&= \overline{(\bar{f}u_{\otimes}^{r(-1)})(g \otimes 1)a_{\otimes}} && \text{Lemma 2.2.3} \\
&= \overline{\bar{f}gu_{\otimes}^{r(-1)}a_{\otimes}} && u_{\otimes}^{r(-1)} \text{ natural} \\
&= \overline{\bar{f}g} && u_{\otimes}^{r(-1)}, a_{\otimes} \text{ iso, Lemma 2.2.3} \\
&= \bar{f}\bar{g} && [\mathbf{R.3}] \text{ in } \mathbb{X}.
\end{aligned}$$

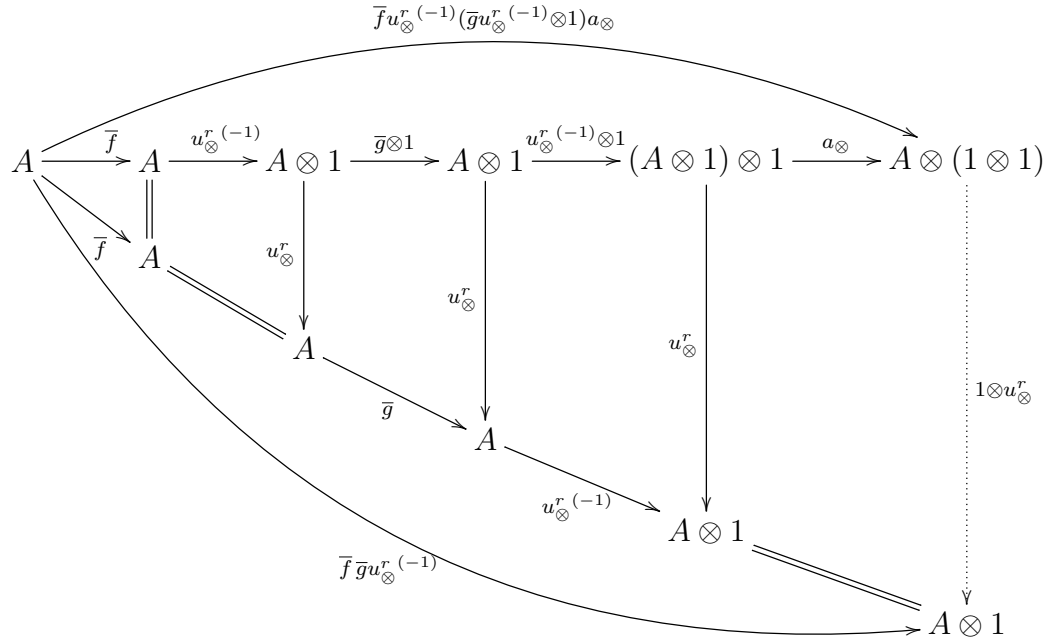
The restriction of the right hand side computes in \mathbb{X} as:

$$\begin{aligned}
\overline{(\bar{f}u_{\otimes}^{r(-1)})(\bar{g}u_{\otimes}^{r(-1)} \otimes 1)a_{\otimes}} &= \overline{(\bar{f}u_{\otimes}^{r(-1)})(\bar{g}u_{\otimes}^{r(-1)} \otimes 1)} && a_{\otimes} \text{ iso, Lemma 2.2.3} \\
&= \overline{\bar{f}\bar{g}u_{\otimes}^{r(-1)}u_{\otimes}^{r(-1)}} && u_{\otimes}^{r(-1)} \text{ natural} \\
&= \overline{\bar{f}\bar{g}} && u_{\otimes}^{r(-1)}u_{\otimes}^{r(-1)} \text{ iso, Lemma 2.2.3} \\
&= \bar{f}\bar{g} && \text{Lemma 2.2.3.}
\end{aligned}$$

Additionally, we see $\overline{\bar{f}g}$ in $\widetilde{\mathbb{X}}$ is expressed in \mathbb{X} as:

$$\begin{aligned}
\overline{(\bar{f}u_{\otimes}^{r(-1)})(g \otimes 1)a_{\otimes}u_{\otimes}^{r(-1)}} &= \bar{f}u_{\otimes}^{r(-1)}\overline{g \otimes 1} && [\mathbf{R.3}], [\mathbf{R.4}], a_{\otimes} \text{ iso} \\
&= \bar{f}\bar{g}u_{\otimes}^{r(-1)} && \otimes \text{a restriction bi-functor, } u_{\otimes}^{r(-1)} \text{ natural.}
\end{aligned}$$

The following diagram in \mathbb{X} follows the above right hand side with the top curved arrow and the left hand side with the bottom curved arrow. Note that we are using that $\overline{(\bar{f}u_{\otimes}^{r(-1)})(g \otimes 1)a_{\otimes}} = \bar{f}\bar{g}$ as shown above.



Hence, in \mathbb{X} , $\overline{(\bar{f}u_{\otimes}^{r(-1)})(g \otimes 1)a_{\otimes}u_{\otimes}^{r(-1)}} \stackrel{1 \otimes u_{\otimes}^r}{\simeq} \overline{(\bar{f}u_{\otimes}^{r(-1)})(\bar{g}u_{\otimes}^{r(-1)} \otimes 1)a_{\otimes}}$ and therefore $\overline{\bar{f}g} = \bar{f}\bar{g}$ in $\tilde{\mathbb{X}}$.

R.4 $\bar{f}\bar{g} = \overline{\bar{f}g}f$ The restriction of the left hand side is:

$$\begin{aligned}
 \overline{f(\bar{g}u_{\otimes}^{r(-1)} \otimes 1)a_{\otimes}} &= \overline{f(\bar{g}u_{\otimes}^{r(-1)} \otimes 1)} && a_{\otimes} \text{ iso, Lemma 2.2.3} \\
 &= \overline{f\bar{g}u_{\otimes}^{r(-1)}} \otimes \bar{f} && \otimes \text{ restriction functor} \\
 &= \overline{f\bar{g}} \otimes \bar{f} && u_{\otimes}^{r(-1)} \text{ iso, Lemma 2.2.3} \\
 &= \overline{f(\bar{g} \otimes 1)}
 \end{aligned}$$

and the restriction of the right hand side is:

$$\begin{aligned}
\overline{f(g \otimes 1)u_{\otimes}^{r(-1)}(f \otimes 1)a_{\otimes}} &= \overline{f(g \otimes 1)u_{\otimes}^{r(-1)}(f \otimes 1)} && a_{\otimes} \text{ iso, Lemma 2.2.3} \\
&= \overline{f(g \otimes 1)fu_{\otimes}^{r(-1)}} && u_{\otimes}^{r(-1)} \text{ natural} \\
&= \overline{f(\bar{g} \otimes 1)u_{\otimes}^{r(-1)}} && [\mathbf{R.4}] \text{ for } \mathbb{X} \\
&= \overline{f(\bar{g} \otimes 1)u_{\otimes}^{r(-1)}} && \otimes \text{ is a restriction functor} \\
&= \overline{f(\bar{g} \otimes 1)} && u_{\otimes}^{r(-1)} \text{ iso, Lemma 2.2.3}
\end{aligned}$$

Computing the right hand side in \mathbb{X} ,

$$\begin{aligned}
\overline{f(g \otimes 1)a_{\otimes}u_{\otimes}^{r(-1)}(f \otimes 1)a_{\otimes}} &= \overline{f(g \otimes 1)fu_{\otimes}^{r(-1)}a_{\otimes}} && a_{\otimes} \text{ iso, } u_{\otimes}^{r(-1)} \text{ natural.} \\
&= \overline{f(\bar{g} \otimes 1)u_{\otimes}^{r(-1)}a_{\otimes}} && [\mathbf{R.3}], \otimes \text{ a restriction functor.}
\end{aligned}$$

$$\begin{array}{ccccccc}
A & \xrightarrow{f} & B \otimes C & \xrightarrow{\bar{g}u_{\otimes}^{r(-1)} \otimes 1} & (B \otimes 1) \otimes C & \xrightarrow{a_{\otimes}} & B \otimes (1 \otimes C) \\
& \searrow f & & & & & \downarrow 1 \otimes c_{\otimes} \\
& & B \otimes C & \xrightarrow{\bar{g} \otimes 1} & B \otimes C & \xrightarrow{u_{\otimes}^{r(-1)}} & (B \otimes C) \otimes 1 \xrightarrow{a_{\otimes}} B \otimes (C \otimes 1)
\end{array}$$

and hence, $\tilde{\mathbb{X}}$ is a restriction category. \square

3.2.2 The category $\tilde{\mathbb{X}}$ is a discrete restriction category

Lemma 3.2.8. *The unit of the inverse product in \mathbb{X} is the terminal object in $\tilde{\mathbb{X}}$.*

Proof. The unique map to the terminal object for any object A in $\tilde{\mathbb{X}}$ is the equivalence class of maps represented by $(u_{\otimes}^{l(-1)}, A)$. For this to be a terminal object, the diagram

$$\begin{array}{ccccc}
X & \xrightarrow{\overline{(f,C)}} & X & \xrightarrow{!_X} & \top \\
\downarrow (f,C) & & & \nearrow !_Y & \\
Y & & & &
\end{array}$$

must commute for all choices of f . Translating this to \mathbb{X} , this is the same as requiring

$$\begin{array}{ccccccc}
X & \xrightarrow{\bar{f}} & X & \xrightarrow{u_{\otimes}^{r(-1)}} & X \otimes 1 & \xrightarrow{u_{\otimes}^{l(-1)}} & 1 \otimes X \otimes 1 \\
\downarrow f & & & & & \swarrow 1 \otimes (u_{\otimes}^r f) & \\
Y \otimes C & \xrightarrow{u_{\otimes}^{l(-1)}} & 1 \otimes Y \otimes C & & & &
\end{array}$$

commute, which is true by [R.1] and from the coherence diagrams for the inverse product tensor. \square

Next, we show that the category $\widetilde{\mathbb{X}}$ has restriction products, given by the action of $(\widetilde{-})$ on the \otimes tensor in \mathbb{X} .

First, define total maps π_0, π_1 in $\widetilde{\mathbb{X}}$ by:

$$\pi_0 : A \otimes B \xrightarrow{(1, B)} A \quad (3.5)$$

$$\pi_1 : A \otimes B \xrightarrow{(c_\otimes, A)} B \quad (3.6)$$

Given the maps $Z \xrightarrow{(f, C)} A$ and $Z \xrightarrow{(g, C')} B$, define $\langle (f, C), (g, C') \rangle$ as

$$Z \xrightarrow{(\Delta(f \otimes g)(1 \otimes c_\otimes \otimes 1), C \otimes C')} A \otimes B \quad (3.7)$$

where associativity is assumed as needed. Note that with the associativity maps, this is actually:

$$Z \xrightarrow{(\Delta(f \otimes g)a_\otimes(1 \otimes a_\otimes^{(-1)})(1 \otimes (c_\otimes \otimes 1))(1 \otimes a_\otimes)a_\otimes^{(-1)}, C \otimes C')} A \otimes B \quad (3.8)$$

Lemma 3.2.9. *On $\widetilde{\mathbb{X}}$, \otimes is a restriction product with projections π_0, π_1 with the product of maps f, g being $\langle f, g \rangle$.*

Proof. From the definition above, as 1 and c_\otimes are isomorphisms, the maps π_0, π_1 are total.

In order to show that $\overline{\langle f, g \rangle} = \overline{f} \overline{g}$, first reduce the left hand side:

$$\begin{aligned} \overline{\langle f, g \rangle} &= \overline{\Delta(f \otimes g)(1 \otimes c_\otimes \otimes 1)u_\otimes^{r(-1)}} && \text{in } \mathbb{X}, \text{ definition of restriction} \\ &= \overline{\Delta(f \otimes g)u_\otimes^{r(-1)}} && c_\otimes \text{ is iso} \\ &= \overline{\Delta(\overline{f} \otimes \overline{g})u_\otimes^{r(-1)}} && \text{from Lemma 2.2.3} \\ &= \overline{\Delta(\overline{f} \otimes \overline{g})u_\otimes^{r(-1)}} && \otimes \text{ is a restriction functor} \\ &= \overline{\overline{f} \overline{g} \Delta(1 \otimes 1)u_\otimes^{r(-1)}} && \text{Lemma 3.1.8(ii) twice} \\ &= \overline{\overline{f} \overline{g} u_\otimes^{r(-1)}} && \text{Lemma 2.2.3} \\ &= \overline{f} \overline{g} u_\otimes^{r(-1)} && \text{Lemma 2.2.3.} \end{aligned}$$

Then, the right hand side reduces as:

$$\begin{aligned}\overline{f\bar{g}} &= \overline{f}u_{\otimes}^r{}^{(-1)}(\overline{g}u_{\otimes}^r{}^{(-1)} \otimes 1)a_{\otimes} && \text{in } \mathbb{X} \text{ by definitions} \\ &= \overline{f\bar{g}}u_{\otimes}^r{}^{(-1)}u_{\otimes}^r{}^{(-1)}a_{\otimes} && u_{\otimes}^r{}^{(-1)} \text{ natural.}\end{aligned}$$

The restriction of the left hand side and the right hand side, in \mathbb{X} , is $\overline{f\bar{g}}$. This is done by applying Lemma 2.2.3 on page 14 once on the left and thrice on the right.

Thus, this shows $\overline{\langle f, g \rangle} = \overline{f\bar{g}}$ in $\widetilde{\mathbb{X}}$ where the mediating map in \mathbb{X} is $1 \otimes u_{\otimes}^r$.

Next, to show $\langle f, g \rangle \pi_0 \leq f$ (and $\langle f, g \rangle \pi_1 \leq g$), it is required to show $\overline{\langle f, g \rangle \pi_0} f = \langle f, g \rangle \pi_0$.

Calculating the left side, we see:

$$\begin{aligned}\overline{\langle f, g \rangle \pi_0} f &= \overline{\langle f, g \rangle \pi_0} f && \text{Lemma 2.2.3} \\ &= \overline{\langle f, g \rangle} f && \pi_0 \text{ is total} \\ &= \overline{f\bar{g}} f && \text{by above} \\ &= \overline{g} \overline{f} f && [\mathbf{R.2}] \\ &= \overline{g} f && [\mathbf{R.1}].\end{aligned}$$

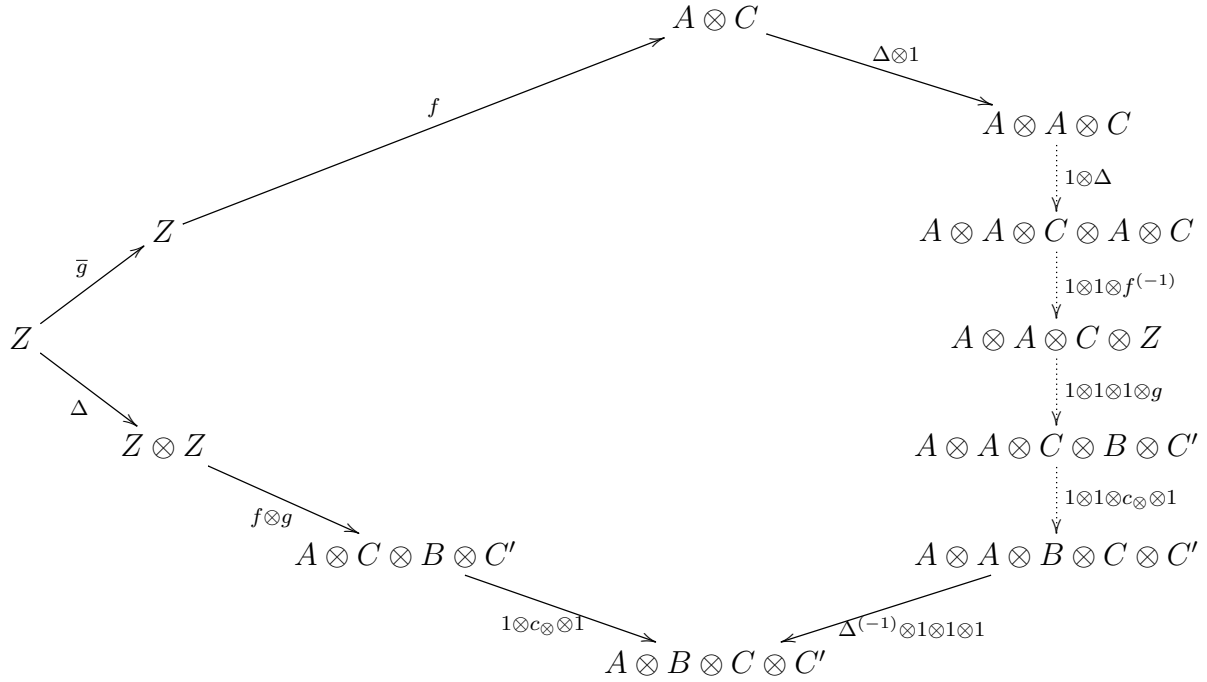
Now, turning to the right hand side:

$$\langle f, g \rangle \pi_0 = \Delta(f \otimes g)(1 \otimes c_{\otimes} \otimes 1)1 \quad \text{in } \mathbb{X}, \text{ by definition.}$$

To show these are equal in $\widetilde{\mathbb{X}}$, we need to first show the restrictions are the same in \mathbb{X} and then show there is a mediating map between the images in \mathbb{X} . The restriction of $\overline{g}f$ is $\overline{f\bar{g}}$ immediately by $[\mathbf{R.3}]$ and $[\mathbf{R.2}]$. For the right hand side, calculate in \mathbb{X} :

$$\begin{aligned}\overline{\Delta(f \otimes g)(1 \otimes c_{\otimes} \otimes 1)} &= \overline{\Delta(f \otimes g)} && \text{Lemma 2.2.3} \\ &= \Delta(f \otimes g)(f^{(-1)} \otimes g^{(-1)})\Delta^{(-1)} && \mathbb{X} \text{ is an inverse category} \\ &= \Delta(\overline{f} \otimes \overline{g})\Delta^{(-1)} \\ &= \overline{f\bar{g}}\Delta\Delta^{(-1)} && \text{Lemma 3.1.8(ii) twice} \\ &= \overline{f\bar{g}}.\end{aligned}$$

The diagram below, shows the required mediating map.



□

At this point, we have shown that $\tilde{\mathbb{X}}$ is a restriction category with restriction products. This leads us to the following theorem:

Theorem 3.2.10. *For any inverse category \mathbb{X} , the category $\tilde{\mathbb{X}}$ is a discrete restriction category.*

Proof. The fact that $\tilde{\mathbb{X}}$ is a Cartesian restriction category is immediate from lemmas 3.2.6 on page 49, 3.2.7 on page 52, 3.2.8 on page 55 and 3.2.9 on page 56.

To show that it is discrete, we need only show that the map $(\Delta u_{\otimes}^{r(-1)}, 1)$ is in the same equivalence class as $\tilde{\mathbb{X}}$'s $\Delta (= \langle 1, 1 \rangle = \langle (u_{\otimes}^{r(-1)}, 1), (u_{\otimes}^{r(-1)}, 1) \rangle)$. As both Δ and $u_{\otimes}^{r(-1)}$ are total, the restriction of each side is the same, namely 1. The diagram below uses Corollary

3.2.5 and shows that the two maps are in the same equivalence class.

$$\begin{array}{ccc}
 & & A \otimes A \otimes 1 \\
 & \nearrow \Delta u_{\otimes}^{r(-1)} & \downarrow u_{\otimes}^{r(-1)} \\
 A & \xrightarrow{\Delta(u_{\otimes}^{r(-1)} \otimes u_{\otimes}^{r(-1)})(1 \otimes c_{\otimes} \otimes 1)} & A \otimes A \otimes 1 \otimes 1
 \end{array}$$

□

3.2.3 Equivalence of categories

This section will show that the category of discrete inverse categories (maps being restriction functors that preserve the inverse tensor) is equivalent to the category of discrete restriction categories (maps being the restriction functors which preserve the product). In the following, \mathbb{X} will always be a discrete inverse category, \mathbb{D} and \mathbb{C} will be discrete restriction categories.

We approach the equivalence proof by exhibiting the universal property for discrete inverse categories for the functor **INV** from discrete restriction categories to discrete inverse categories. The functor **INV** maps a discrete restriction category to its inverse subcategory and maps functors between discrete restriction categories to a functor having the same action on the partial inverses. That is, given $G : \mathbb{C} \rightarrow \mathbb{D}$, then:

$$\mathbf{INV}(G) : \mathbf{INV}(\mathbb{C}) \rightarrow \mathbf{INV}(\mathbb{D})$$

$$\mathbf{INV}(G)(A) = GA \quad (\text{all objects of } \mathbb{D} \text{ are in } \text{Inv}(\mathbb{D}))$$

$$\mathbf{INV}(G)(f) = G(f) \quad (\text{restriction functors preserve partial inverse})$$

We continue by showing the η and ε of the universal property are isomorphisms. First, let $\eta : \mathbb{X} \rightarrow \mathbf{INV}(\tilde{\mathbb{X}})$ be an identity on objects functor. For maps f in \mathbb{X} , $\eta(f) = (fu_{\otimes}^{r(-1)}, 1)$.

Next, consider a functor $F : \mathbb{X} \rightarrow \mathbf{INV}(\mathbb{D})$ defined as follows:

$$\text{Objects: } F^{\#} : A \mapsto F(A)$$

$$\text{Arrows: } F^{\#} : (f, C) \mapsto F(f)\pi_0$$

This allows us to write the diagram:

$$\begin{array}{ccc}
 \mathbb{X} & \xrightarrow{\eta} & \mathbf{INV}(\tilde{\mathbb{X}}) \\
 & \searrow F & \downarrow \mathbf{INV}(F^\#) \\
 & & \mathbf{INV}(\mathbb{D})
 \end{array} \tag{3.9}$$

In order to show this is a universal diagram, we proceed with a series of lemmas building to the result.

Lemma 3.2.11. *For any discrete inverse category \mathbb{X} , all invertible maps $(g, C) : A \rightarrow B$ in $\tilde{\mathbb{X}}$ are in the equivalence class of $(fu_\otimes^r{}^{(-1)}, 1)$ for some $f : A \rightarrow B$.*

Proof. As (g, C) is invertible in $\tilde{\mathbb{X}}$, the map $(g, C)^{(-1)} : B \rightarrow A$ exists. $(g, C)^{(-1)}$ must be in the equivalence class of some map $k : B \rightarrow A \otimes D$, and also note that $\overline{(g, C)}$ is by construction the equivalence class of the map $\bar{g}u_\otimes^r{}^{(-1)} : A \rightarrow A \otimes 1$ in \mathbb{X} . This means, diagramming in \mathbb{X} , there is an n such that

$$\begin{array}{ccccc}
 B & \xrightarrow{k} & A \otimes D & \xrightarrow{f \otimes 1} & B \otimes C \otimes D \\
 & & & & \downarrow \Delta \otimes 1 \\
 & & & & B \otimes B \otimes C \otimes D \\
 & & & & \vdots 1 \otimes n \\
 & & & & B \otimes B \otimes 1 \\
 & & & & \downarrow \Delta^{(-1)} \otimes 1 \\
 & & & & B \otimes 1 \\
 & \searrow \bar{g}u_\otimes^r{}^{(-1)} & & &
 \end{array}$$

commutes.

Starting with $g : A \rightarrow B \otimes C$, construct the map f in \mathbb{X} with the following diagram:

$$\begin{array}{ccc}
 A & \xrightarrow{g} & B \otimes C \\
 & \searrow f & \downarrow \Delta \otimes 1 \\
 & & B \otimes B \otimes C \\
 & & \downarrow 1 \otimes \Delta \otimes 1 \\
 & & B \otimes B \otimes B \otimes C \\
 & & \downarrow 1 \otimes 1 \otimes k \otimes 1 \\
 & & B \otimes B \otimes A \otimes D \otimes C \\
 & & \downarrow 1 \otimes 1 \otimes g \otimes 1 \otimes 1 \\
 & & B \otimes B \otimes B \otimes C \otimes D \otimes C \\
 & & \downarrow 1 \otimes \Delta^{(-1)} \otimes 1 \otimes c_{\otimes} \\
 & & B \otimes B \otimes C \otimes C \otimes D \\
 & & \downarrow 1 \otimes 1 \otimes \Delta^{(-1)} \otimes 1 \\
 & & B \otimes B \otimes C \otimes D \\
 & & \downarrow 1 \otimes n \\
 & & B \otimes B \otimes 1 \\
 & & \downarrow (\Delta^{(-1)} \otimes 1) u_{\otimes}^l \\
 & & B
 \end{array}$$

By its construction, $f : A \rightarrow B$ in \mathbb{X} and $(fu_{\otimes}^{r(-1)}, 1)$ is in the same equivalence class as (g, C) .

□

Lemma 3.2.12. *Diagram (equation (3.9)) above is a commutative diagram.*

Proof. Chasing maps around the diagram, we have:

$$\begin{array}{ccc}
 f & \xrightarrow{\eta} & (fu_{\otimes}^{r(-1)}, 1) \\
 & \searrow F & \downarrow \mathbf{INV}(F\#) \\
 & & F(f) \equiv F(fu_{\otimes}^{r(-1)})\pi_0
 \end{array}$$

As η is identity on the objects, diagram equation (3.9) on the preceding page commutes. □

Lemma 3.2.13. *The functor \mathbf{INV} from the category of discrete restriction categories to the category of discrete inverse categories is full and faithful.*

Proof. To show fullness, we must show **INV** is surjective on hom-sets. Given a functor between two categories in the image of **INV**, i.e., $G : \mathbf{INV}(\mathbb{C}) \rightarrow \mathbf{INV}(\mathbb{D})$, construct a functor $H : \mathbb{C} \rightarrow \mathbb{D}$ as follows:

Action on objects: $H(A) = G(A)$,

Objects on maps: $H(f) = G(\langle f, 1 \rangle)\pi_0$.

H is well defined as we know $\langle f, 1 \rangle$ is an invertible map and therefore in the domain of G .

To see H is a functor:

$$H(1) = G(\langle 1, 1 \rangle)\pi_0 = \Delta_{\mathbb{D}}\pi_0 = 1$$

$$H(fg) = G(\langle fg, 1 \rangle)\pi_0 = G(\langle f, 1 \rangle)\pi_0 G(\langle g, 1 \rangle)\pi_0 = H(f)H(g)$$

But on any invertible map, $H(f) = G(\langle f, 1 \rangle)\pi_0 = \langle G(f), 1 \rangle\pi_0 = G(f)$ and therefore $\mathbf{INV}((\)H) = G$, so **INV** is full.

Next, assume we have $F, G : \mathbb{C} \rightarrow \mathbb{D}$ with $\mathbf{INV}(F) = \mathbf{INV}(G)$. Considering $F(f)$ and $F(g)$, we know $F(\langle f, 1 \rangle) = G(\langle f, 1 \rangle)$ as $\langle f, 1 \rangle$ is invertible. Thus, as the functors preserve the product structure, we have

$$F(f) = F(\langle f, 1 \rangle)F(\pi_0) = G(\langle f, 1 \rangle)G(\pi_0) = G(f).$$

Thus, **INV** is faithful. □

Corollary 3.2.14. *The functor $F^\#$ in diagram [equation \(3.9\) on page 60](#) is unique.*

Proof. This follows immediately from lemma [3.2.13 on the preceding page](#), **INV** is faithful. □

Corollary 3.2.15. *The category $\tilde{\mathbb{X}}$ and functor $\eta : \mathbb{X} \rightarrow \mathbf{INV}(\tilde{\mathbb{X}})$ is a universal pair for the functor **INV**.*

Proof. Immediate from Corollary [3.2.14](#) and Lemma [3.2.12 on the preceding page](#). □

Lemma 3.2.16. *The functor $\eta : \mathbb{X} \rightarrow \mathbf{INV}(\widetilde{\mathbb{X}})$ is an isomorphism.*

Proof. As η is an identity on objects functor, we need only show that it is full and faithful. Referring to Lemma 3.2.11 on page 60 above, we immediately see that η is full. For faithful, if we assume $(fu_{\otimes}^{r(-1)}, 1)$ is equal in $\widetilde{\mathbb{X}}$ to $(gu_{\otimes}^{r(-1)}, 1)$. This means in \mathbb{X} , that $\bar{f} = \bar{g}$ and there is a h such that

$$\begin{array}{ccccc}
 & & B \otimes 1 & \xrightarrow{(\Delta \otimes 1) a_{\otimes}} & B \otimes (B \otimes 1) \\
 & \nearrow fu_{\otimes}^{r(-1)} & & & \downarrow 1 \otimes h \\
 A & & & & B \otimes (B \otimes 1) \\
 & \searrow gu_{\otimes}^{r(-1)} & & & \uparrow a_{\otimes}^{(-1)} (\Delta^{(-1)} \otimes 1) \\
 & & B \otimes 1 & &
 \end{array}$$

This simplifies out to $g = f\Delta(1 \otimes h)\Delta^{(-1)}$. But by Lemma 3.1.8 on page 38, part (iv) on page 38, $\Delta(1 \otimes h)\Delta^{(-1)} = \overline{\Delta(1 \otimes h)\Delta^{(-1)}}$. Setting $\Delta(1 \otimes h)\Delta^{(-1)}$ as k , we have $g = f\bar{k}$. But this gives us:

$$g = f\bar{k} = \overline{f\bar{k}}f = \overline{f\bar{k}}f = \bar{g}f = \bar{f}f = f.$$

This shows η is faithful and hence an isomorphism between \mathbb{X} and $\mathbf{INV}(\widetilde{\mathbb{X}})$. \square

Theorem 3.2.17. *The category of discrete inverse categories (objects are discrete inverse categories, maps are inverse tensor preserving functors) is equivalent to the category of discrete restriction categories (objects are discrete restriction categories, maps are the Cartesian restriction functors).*

Proof. From the above lemmas, we have shown that we have an adjoint:

$$(\eta, \varepsilon) : \mathbf{T} \vdash \mathbf{INV} : D_{ic} \rightarrow D_{rc} \quad (3.10)$$

By lemma 3.2.16 we know η is an isomorphism. But this means the functor \mathbf{T} is full and faithful, as shown in, e.g., Proposition 2.2.6 of [19]. From lemma 3.2.13 we know that \mathbf{INV}

is full and faithful. But again by the previous reference, this means ε is an isomorphism. Thus, by Corollary 3.2.15 and Proposition 2.2.7 of [19] we have the equivalence of the two categories. \square

3.2.4 Examples of the $\widetilde{(-)}$ construction

Example 3.2.18 (Completing a finite discrete inverse category).

Continuing from example 3.1.7 on page 37, recall the discrete category of 4 elements with two different tensors. Completing these gives two different lattices. They are either the straight line lattice, or the diamond semilattice. Below are the details of these constructions.

Recall \mathbb{D} has four elements a, b, c and d , and there are two possible inverse product tensors:

\otimes	a	b	c	d
a	a	a	a	a
b	a	b	b	b
c	a	b	c	c
d	a	b	c	d

\otimes	a	b	c	d
a	a	a	a	a
b	a	b	a	b
c	a	a	c	c
d	a	b	c	d

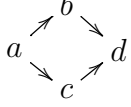
Define Δ as the identity map. Then, for the first tensor, $\widetilde{\mathbb{D}}$ has the following maps

$$\begin{array}{llll}
 a \xrightarrow{(id,a) \ (\equiv(id,b) \equiv(id,c) \equiv(id,d))} a, & a \xrightarrow{(id,a)} b, & a \xrightarrow{(id,a)} c, & a \xrightarrow{(id,a)} d \\
 b \xrightarrow{(id,b) \ (\equiv(id,c) \equiv(id,d))} b, & b \xrightarrow{(id,b)} c, & b \xrightarrow{(id,b)} d & \\
 c \xrightarrow{(id,c) \ (\equiv(id,d))} c, & c \xrightarrow{(id,c)} d & & \\
 d \xrightarrow{(id,d)} d & & &
 \end{array}$$

resulting in the straight-line $(a \rightarrow b \rightarrow c \rightarrow d)$ lattice. The tensor in \mathbb{D} becomes the meet and hence is a categorical product in $\widetilde{\mathbb{D}}$. Note that the only partial inverses in $\widetilde{\mathbb{D}}$ are the identity functions and that for all maps f , $\langle f, 1 \rangle = id$.

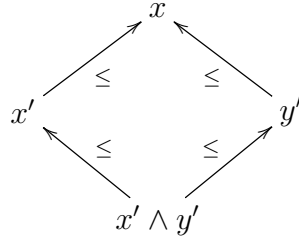
With the second tensor table, we have:

$$\begin{array}{llll}
a \xrightarrow{(id,a) \equiv (id,b) \equiv (id,c) \equiv (id,d)} a, & a \xrightarrow{(id,a)} b, & a \xrightarrow{(id,a)} c, & a \xrightarrow{(id,a)} d \\
& b \xrightarrow{(id,b) \equiv (id,d)} b, & b \xrightarrow{(id,b)} d & \\
& c \xrightarrow{(id,c) \equiv (id,d)} c, & c \xrightarrow{(id,c)} d & \\
& & d \xrightarrow{(id,d)} d &
\end{array}$$

resulting in the “diamond” lattice, . Once again, the tensor in \mathbb{D} is the meet.

Example 3.2.19. Lattice completion. Suppose we have a set together with an idempotent, commutative, associative operation \wedge on the set, giving us a lattice, \mathbb{L} . Further suppose the set is partially ordered via \leq with the order being compatible with \wedge .

Then, we may create a pullback square for any $x' \leq x$, $y' \leq x$ with



Considering \mathbb{L} as a category, we see that all maps are monic and therefore, we may create a partial map category $\text{Par}(\mathbb{L}, \mathcal{M})$ where the stable system of monics are all the maps.

Then $\widetilde{\text{Par}(\mathbb{L}, \mathcal{M})}$ becomes the completion of the lattice over \wedge .

3.3 Reversible computation

Bennet, in [7], showed that it was possible to emulate a standard Turing machine via a reversible Turing machine and vice-versa. This showed the equivalence of standard and reversible Turing machines. We reproduce the essence of this proof below.

3.3.1 Reversible Turing machines

Turing machines consist of a tape, a read-write head positioned over the tape, a machine state and a set of instructions. The set of instructions may be given as a set of transitions determining the movement of the read-write head, what it writes and the resulting state of the machine.

Definition 3.3.1. Given an alphabet A which does not contain a space, a tape is in *standard format* when:

- [T.1] The tape head is positioned directly over a blank space;
- [T.2] The spaces to the left (the $+1$ direction) contain only elements of A .
- [T.3] All other spaces of the tape are blank.

Definition 3.3.2. A *turing quintuple* is a quintuple $(s, \alpha, \alpha', \delta, s')$ where:

- [Q.1] $s, s' \in S$, where S is a predefined set of states;
- [Q.2] $\alpha, \alpha' \in A$ is predefined set of glyphs;
- [Q.3] $\delta \in \{-1, 0, 1\}$.

Definition 3.3.3. A *standard turing quintuple set* Q consists of a set of turing quintuples such that:

- (i) If $q_1 = (s_1, \alpha_1, \alpha'_1, \delta_1, s'_1)$ and $q_2 = (s_2, \alpha_2, \alpha'_2, \delta_2, s'_2)$ are in Q , then either $s_1 \neq s_2$ or $\alpha_1 \neq \alpha_2$ or both are not equal.

- (ii) There are two special quintuples contained in Q :

(a) $(s_1, \sqcup, \sqcup, +1, s_2)$ ¹, the *start quintuple*;

(b) $(s_{t-1}, \sqcup, \sqcup, 0, s_t)$, the *end quintuple* where t is the number of states and is the final state of the machine.

¹Here, \sqcup is used to signify a blank.

Definition 3.3.4. A *standard Turing machine* is given by

- [TM.1] a standard turing quintuple set;
- [TM.2] a tape that starts in standard format;
- [TM.3] and the condition that and if the machine halts, it will halt in state s_t , the final state of the end quintuple and the output will be in standard format.

The turing quintuples may also be regarded as giving the data for a partial function in SETS: $\tau : S \times A \rightarrow A \times \{-1, 0, 1\} \times S$.

Remark 3.3.5. A multi-tape Turing machine with n tapes and read-write heads can be described by modifying definition 3.3.4 such that α is an n -tuple of the set of glyphs for the Turing machine and δ is an n -tuple of movement directions.

Example 3.3.6. Suppose $S = \{start, run, reset, done\}$, $A = \{0, 1, \sqcup\}$ and the Turing machine program is given by the quintuples

$$\begin{aligned}
 &(start, \sqcup, \sqcup, +1, run), \\
 &(run, 0, 1, +1, run), (run, 1, 0, +1, run), \\
 &(run, \sqcup, \sqcup, -1, reset), \\
 &(reset, 0, 0, -1, reset), (reset, 1, 1, -1, reset), \\
 &(reset, \sqcup, \sqcup, 0, done).
 \end{aligned}$$

This program will perform a “bit-flip” of all 0s and 1s on the tape until it reads a space, reposition the read head to the standard format and then it will halt.

As we see in example 3.3.6, it is *possible* that a Turing machine program is reversible. If we had chosen the second quintuple to be $(run, 0, 0, +1, run)$ instead, the program would not have been reversible.

The essential property that a Turing machine program needs to be reversible is that the function τ defined from the quintuples is injective. In order to simplify the discovery the function being injective, we reformulate the turing quintuples as quadruples.

Definition 3.3.7. A *turing quadruple* is given by a quadruple

$$(s, [b_1, b_2, \dots, b_n], [b'_1, b'_2, \dots, b'_n], s')$$

such that:

- (i) $s, s' \in S$, some set of states;
- (i) $b_j \in A \cup \{\phi\}$ where A is some alphabet;
- (i) $b'_j \in A + \{-1, 0, 1\}$;
- (i) $b'_j \in \{-1, 0, 1\}$ if and only if $b_j = \phi$.

In this definition, $b_j = \phi$ means that the value of tape j is ignored.

A turing quadruple explicitly splits the read/write action of the Turing machine away from the movement. In a particular step for tape k , the turing machine will either read and write an item or it will move.

Remark 3.3.8. Any turing quintuple q of n tapes may be split into two turing quadruples, q_r and q_m by the addition of a new state a'' in A . The quadruple q_r will consist of all the read-write operations and leave the Turing machine in state a'' . The quadruple q_m will start in state a'' with all the b_j set to ϕ and b'_j being movement on each of the n tapes.

Definition 3.3.9. A set of turing quadruples Q is called *reversible set of turing quadruples* when given $q_1, q_2 \in Q$, with $q_1 = (a, [b_j], [b'_j], a')$ and $q_2 = (c, [d_j], [d'_j], c')$:

[RTM.1] if $a = c$, then there is a k where $b_k, d_k \in A$ and $b_k \neq d_k$;

[RTM.2] if $a' = c'$, then there is a j with $b'_j, d'_j \in A$ and $b'_j \neq d'_j$.

Similarly to turing quintuples, turing quadruples may be taken as the data for a function in SETS:

$$\rho : S \times (A \cup \{\phi\}) \rightarrow (A + \{-1, 0, 1\}) \times S.$$

We can see by inspection that ρ is a reversible partial function when the set of turing quadruples that give ρ is a reversible set of turing quadruples.

Definition 3.3.10. A *reversible Turing machine* is one that is described by a set of reversible turing quadruples.

We will show that a reversible Turing machine with three tapes can emulate a Turing machine.

Theorem 3.3.11 (Bennet[7]). *Given a standard Turing Machine M , it may be emulated by a three tape reversible Turing machine R . In this case, emulated means:*

- (i) M halts on standard input I if and only if R halts on standard input (I, \sqcup, \sqcup) .
- (i) M halts on standard input I producing standard output O , if and only if R halts on input (I, \sqcup, \sqcup) producing standard output (I, \sqcup, O) .

Proof. (Sketch only).

The crux of the proof is to convert the quintuples of M to the quadruples of R as noted in remark 3.3.8 on the preceding page. Explicitly for a single tape machine, we have

$$(s, a, a, \delta, s') \mapsto ((s, a, a', s''), (s'', \phi, \delta, s')). \quad (3.11)$$

In equation (3.11), s'' is a new state for the machine M , not in the current set of states.

Assign an order to the n quintuples of M , where the start quintuple is the first in the order and the end quintuple comes last. Convert these to quadruples as in equation (3.11).

We then proceed to create three groups of quadruples for R . We call these *emulation*, *copy*, and *restore*.

To create the emulation phase quadruples, we examine the pairs of quadruples of M in the sorted order and produce a pair of quadruples for R .

$$\begin{aligned}
\text{Pair 1} \quad & (s_1, \sqcup, \sqcup, s_1'') \mapsto (s_1, [\sqcup, \phi, \sqcup], [\sqcup, +1, \sqcup], e_1) \\
& (s_1'', \phi, \delta, s_2) \mapsto (e_1, [\phi, \sqcup, \phi], [\delta, 1, 0], s_2) \\
& \vdots \\
\text{Pair } j \quad & (s_k, a_j, a_j', s_k'') \mapsto (s_k, [a_j, \phi, \sqcup], [a_j', +1, \sqcup], e_j) \\
& (s_k'', \phi, \delta, s_i) \mapsto (e_j, [\phi, \sqcup, \phi], [\delta_j, j, 0], s_i) \\
& \vdots \\
\text{Pair } n \quad & (s_\ell, \sqcup, \sqcup, s_\ell'') \mapsto (s_\ell, [\sqcup, \phi, \sqcup], [\sqcup, +1, \sqcup], e_n) \\
& (s_\ell'', \phi, 0, s_f) \mapsto (e_n, [\phi, \sqcup, \phi], [0, n, 0], s_f).
\end{aligned}$$

By inspection, one can see that even if the quadruples of M were not a reversible set, the set created for R is a reversible set, due to the writing of the quadruple index on tape 2. Upon completion of the emulation phase, tape 1 will be the same as M would have produced on its single tape, tape 2 will be $[1, 2, \dots, n]$ and tape 3 will be blanks.

For the copy phase, we create the following quadruples:

$$\begin{aligned}
& (s_f, [\sqcup, n, \sqcup], [\sqcup, n, \sqcup], c_1) \\
& (c_1, [\phi, \phi, \phi], [+1, 0, +1], c_1') \\
& (c_1', [x, n, \sqcup], [x, n, x], c_1) \quad \text{when } x \neq \sqcup \\
& (c_1', [\sqcup, n, \sqcup], [\sqcup, n, x], c_2) \\
& (c_2, [\phi, \phi, \phi], [-1, 0, -1], c_2') \\
& (c_2', [x, n, x], [x, n, x], c_2) \quad \text{when } x \neq \sqcup \\
& (c_2', [\sqcup, n, \sqcup], [\sqcup, n, \sqcup], r_\ell).
\end{aligned}$$

In these quadruples, the states $\{c_1, c_1', c_2, c_2'\}$ should be chosen to be distinct from the states

in the emulation phase. As an example, set them as follows:

$$c_1 = (\{c\}, s_1) \quad c'_1 = (\{c'\}, s_1) \quad c_2 = (\{c\}, s_f) \quad c'_1 = (\{c'\}, s_f).$$

At the completion of this phase, tapes 1 and 2 will be unchanged and tape 3 will be a copy of tape 1.

Finally we perform the restore phase where the history will be erased and tape 1 reset to the input. The quadruples that will accomplish this are:

$$\begin{aligned} \text{Pair } n & \quad (r_n, [\phi, n, \phi], [0, \sqcup, 0], r'_n) \\ & \quad (r'_n, [\sqcup, \phi, \sqcup], [\sqcup, -1, \sqcup], r_{n-1}) \\ & \quad \vdots \\ \text{Pair } j & \quad (r_k, [\phi, j, \phi], [-\delta_j, \sqcup, 0], r'_j) \\ & \quad (r'_j, [a'_j, \phi, \sqcup], [a_j, -1, \sqcup], r_i) \\ & \quad \vdots \\ \text{Pair } 1 & \quad (r_2, [\phi, 1, \phi], [-1, \sqcup, 0], r'_1) \\ & \quad (r'_1, [\sqcup, \phi, \sqcup], [\sqcup, -1, \sqcup], r_1). \end{aligned}$$

The r states are derived from the s states of the emulation phase.

$$r_j = (\{r\}, s_j) \quad r'_j = (\{r'\}, s_j).$$

In this restore phase, the indexes of the states r match up to the indexes of states s . The quadruples reverse the actions of the emulate phase on tape 1, erase the history on tape 2 and make no change to tape 3.

□

3.3.2 Reversible automata and linear combinatory algebras

While reversible Turing machines, as described in [sub-section 3.3.1 on page 66](#), show that reversible computing is as powerful as standard computing, they do not give us a sense of what may be considered to be happening at a higher level.

To accomplish that task we examine the results of the paper “A Structural Approach to Reversible Computation”[1]. In this paper, Abramsky gives a description of a reversible automaton together with a linear combinatory algebra. We will begin by revisiting some definitions and constructions necessary for discussing automata. The next subsection will introduce combinatory algebras, after which we will describe the reversible automata of [1] and add a short proof that it can emulate a reversible turing machine.

Automata

We will describe the automata as a term-rewriting system. This requires, of course, giving a few basic definitions. See, e.g., [6].

Definition 3.3.12. An *arity* is a function from a function to the natural numbers. The arity of F is the number of inputs (arguments) required by F .

Definition 3.3.13. A *signature* Σ is a set of *function symbols* F, G, \dots , each of which has an arity.

Remark 3.3.14. We refer to functions with low arity in the following ways:

- *Arity* = 0. These are known as *nullary* functions or constants.
- *Arity* = 1. These are known as *unary* functions.
- *Arity* = 2. These are known as *binary* functions.

Definition 3.3.15. A *term alphabet* is a set A containing a signature Σ and a countably infinite set X , the variables. Furthermore, $\Sigma \cap X = \phi$.

Definition 3.3.16. A *term algebra* of the term alphabet $\Sigma \cup X$ is denoted by $T_\Sigma(X)$ and defined as follows:

- $x \in V \implies x \in T_\Sigma$ and
- For any $F \in \Sigma$, with $\text{arity}(F) = n$, and $\{t_1, \dots, t_n\} \subseteq T_\Sigma$, then $F(t_1, \dots, t_n) \in T_\Sigma$. In the case where $\text{arity}(F) = 0$, we write $F \in T_\Sigma$.

Definition 3.3.17. The *ground terms* of a term algebra are those terms that do not contain any variable. The set of these terms is designated as T_Σ .

Remark 3.3.18. Note the ground terms consist of the constants and recursively applying the function symbols of Σ to them.

As we are considering rewrite systems, we will need to consider aspects of substitution and unification.

Definition 3.3.19. A *substitution* is a map $\sigma : T_\Sigma(X) \rightarrow T_\Sigma(X)$ which is natural for all function symbols in Σ . In particular if $\text{arity}(c) = 0$ then $\sigma(c) = c$.

Note that given the above definition a substitution σ is completely determined by its action on variables. If $\sigma : X \rightarrow X$ and is injective, we call σ a renaming. Moreover, if σ restricted to the variables in a term t is an injective map of X on those variables, we call *sigma* a renaming of t .

Substitution allows us to define a partial order on $T_\Sigma(X)$, as follows:

Definition 3.3.20. In $T_\Sigma(X)$, let $\sigma(t) = s$. Then we say s is an *instance* of t , written $s \preceq t$. Moreover, if σ is not just a renaming for t , then we write $s \prec t$. If σ is a renaming of t , we write $s \simeq t$.

Lemma 3.3.21. *Subsumption, as defined in 3.3.20 is a partial order, i.e., it is transitive and reflexive.*

Proof.

□

Lemma 3.3.22. *Given terms r, t such that there is at least one s with $s \preceq r$ and $s \preceq t$, then there exists a g such that $g \preceq r$ and $g \preceq t$ and for any s' with $s' \preceq r$ and $s' \preceq t$ we will have $s' \preceq g$.*

Proof.

1. Algorithm to compute supremum of p, q terms.
2. Strict \prec has no infinite ascending chains.
3. Shows main part - there exists.
4. Can now show it is unique up to renaming.

□

The subsumption ordering can be used to derive a similar ordering on substitutions:

Definition 3.3.23. $\sigma \preceq \tau$ if and only if there is a ρ with $\sigma = \tau\rho$, where $\tau\rho$ is the diagrammatic order composition of the two substitutions.

Definition 3.3.24. For terms s, t , if $\sigma(t) = \sigma(s)$, then the substitution σ is called a *unifier* for the terms s, t .

Lemma 3.3.25. *If s, t are terms with a unifier σ , there exists a substitution τ that unifies s, t such that $\tau \preceq \rho$ whenever ρ unifies s, t . ρ is called the most general unifier of s and t .*

Proof. Follows from 3.3.22 on the previous page. □

Notation 3.3.26. Following [1], we write $\mathcal{U}(t, u) \downarrow \sigma$ if σ is the most general unifier of terms t, u .

Combinatory Algebra

Definition 3.3.27. A *combinatory algebra* is an algebra with one binary operation, \cdot written in infix notation. The operation is not assumed to be associative. Multi-element expressions such as $a \cdot b \cdot c$ are to be taken as associating to the left, that is,

$$a \cdot b \cdot c = (a \cdot b) \cdot c.$$

The combinatory algebra may possess distinguished elements that are subject to specific rewrite rules.

Definition 3.3.28. *Combinatory logic* is the combinatory algebra with two distinguished elements, K and S , such that the following hold:

$$\begin{aligned} K \cdot x \cdot y &= x \\ S \cdot x \cdot y \cdot z &= x \cdot z \cdot (y \cdot z). \end{aligned}$$

Note that combinatory logic does not require a specific set that must be used for the algebra, simply that it has the two distinguished elements.

Combinatory logic was shown to be equivalent to the λ calculus by

For example, we may define the identity combinator I as $I = S \cdot K \cdot K$. Further combinators may be defined, such as the B combinator, defined by $B \cdot a \cdot b \cdot c = a \cdot (b \cdot c)$. The S and K combinators are complete, in that other combinators such as B may be defined from them. E.g., $B = S \cdot (K \cdot S) \cdot K$. In fact, we may define an alternate combinatory algebra that is equivalent to Combinatory Logic.

Definition 3.3.29. A *BCKW-Combinatory algebra* is a Combinatory Algebra with four distinguish elements, B , C , K , and W subject to the following equations:

$$\begin{aligned} B \cdot a \cdot b \cdot c &= a \cdot (b \cdot c) \\ C \cdot a \cdot b \cdot c &= a \cdot c \cdot b \\ K \cdot a \cdot b &= a \\ W \cdot a \cdot b &= a \cdot b \cdot b \end{aligned}$$

In fact, a BCKW-Combinatory algebra is equivalent to a Combinatory logic.

Lemma 3.3.30. *The distinguished elements of a BCKW-Combinatory algebra may be represented by S and K . Conversely, the S and K of a Combinatory logic may be created from B, C, K and W .*

Proof. For the first statement, we have:

$$\begin{aligned}
B &= S \cdot (K \cdot S) \cdot K \\
C &= S \cdot (S \cdot (K \cdot (S \cdot (K \cdot S) \cdot K)) \cdot S) \cdot (K \cdot K) \\
K &= K \\
W &= S \cdot S \cdot (S \cdot K).
\end{aligned}$$

Going the other direction, we have:

$$\begin{aligned}
I &= W \cdot K \\
K &= K \\
S &= B \cdot (B \cdot (B \cdot W) \cdot C) \cdot (B \cdot B) \text{ and} \\
&= B \cdot (B \cdot W) \cdot (B \cdot B \cdot C).
\end{aligned}$$

We show the computations of B and S in detail.

$$\begin{aligned}
B \cdot a \cdot b \cdot c &= S \cdot (K \cdot S) \cdot K \cdot a \cdot b \cdot c \\
&= (K \cdot S) \cdot a \cdot (K \cdot a) \cdot b \cdot c \\
&= S \cdot (K \cdot a) \cdot b \cdot c \\
&= K \cdot a \cdot c \cdot (b \cdot c) \\
&= a \cdot (b \cdot c)
\end{aligned}$$

$$\begin{aligned}
S \cdot a \cdot b \cdot c &= B \cdot (B \cdot W) \cdot (B \cdot B \cdot C) \cdot a \cdot b \cdot c \\
&= (B \cdot W) \cdot ((B \cdot B \cdot C) \cdot a) \cdot b \cdot c \\
&= B \cdot W \cdot ((B \cdot B \cdot C) \cdot a) \cdot b \cdot c \\
&= W \cdot (((B \cdot B \cdot C) \cdot a) \cdot b) \cdot c \\
&= (((B \cdot B \cdot C) \cdot a) \cdot b) \cdot c \cdot c \\
&= B \cdot B \cdot C \cdot a \cdot b \cdot c \cdot c \\
&= B \cdot (C \cdot a) \cdot b \cdot c \cdot c \\
&= (C \cdot a) \cdot (b \cdot c) \cdot c \\
&= C \cdot a \cdot (b \cdot c) \cdot c \\
&= a \cdot c \cdot (b \cdot c)
\end{aligned}$$

□

If we use the notation $a^n \cdot b$ to mean $a \cdot a \cdot \dots \cdot a \cdot b$ where a is repeated n times, then we can terms which correspond to the Church numbers of lambda calculus:

$$\bar{n} \equiv (S \cdot B)^n \cdot (K \cdot I)$$

Definition 3.3.31. A partial function $f : \mathbb{N} \rightarrow \mathbb{N}$ is *representable* in combinatory logic if there is a term M_f such that $M_f \cdot \bar{n} = \bar{m}$ whenever $f(n) = m$ and $M_f \cdot \bar{n}$ does not have a normal form if $f(n) \uparrow$.

When we say that combinatory logic with S and K is complete, we mean the following theorem:

Theorem 3.3.32. *The partial functions that are representable in combinatory logic are exactly the partial recursive functions.*

Linear Combinatory Algebra

Definition 3.3.33. A *Linear Combinatory Algebra* $(A, \cdot, !)$ is an algebra A with an applicative binary operation \cdot , an unary operator $! : A \rightarrow A$ and eight distinguished elements: B, C, I, K, D, δ , F and W in A which satisfy the following rules:

1. $B \cdot a \cdot b \cdot c = a \cdot (b \cdot c)$
2. $C \cdot a \cdot b \cdot c = a \cdot c \cdot b$
3. $I \cdot a = a$
4. $K \cdot a \cdot !b = a$
5. $D \cdot !a = a$
6. $\delta \cdot !a = !!a$
7. $F \cdot !a \cdot !b = !(a \cdot b)$
8. $W \cdot a \cdot !b = a \cdot !b \cdot !b$

Note that a Linear Combinatory Algebra always contains a BCKW-Combinatory algebra.

Define $D' = C \cdot (B \cdot B \cdot I) \cdot (B \cdot D \cdot I)$ and the binary operator \bullet on A such that $a \bullet b \equiv a \cdot !b$.

Then, define the following:

$$\begin{aligned}
 B_s &= C \cdot (B \cdot (B \cdot B \cdot B) \cdot (D' \cdot I)) \cdot (C \cdot ((B \cdot B) \cdot F) \cdot \delta) \\
 C_s &= D' \cdot C \\
 K_s &= D' \cdot K \\
 W_s &= D' \cdot W.
 \end{aligned}$$

Lemma 3.3.34. *Given and Linear Combinatory Algebra $(A, \cdot, !)$, then (A, \bullet) is a BCKW-Combinatory algebra with B, C, K, W set to B_s, C_s, K_s, W_s from above.*

Proof. We show the calculation for K_s , the others are similar.

$$\begin{aligned}
K_s \bullet a \bullet b &\equiv D' \cdot K \cdot !a \cdot !b \\
&= C \cdot (B \cdot B \cdot I) \cdot (B \cdot D \cdot I) \cdot K \cdot !a \cdot !b \\
&= (B \cdot B \cdot I \cdot K) \cdot (B \cdot D \cdot I) \cdot !a \cdot !b \\
&= B \cdot (I \cdot K) \cdot (B \cdot D \cdot I) \cdot !a \cdot !b \\
&= (I \cdot K) \cdot ((B \cdot D \cdot I) \cdot !a) \cdot !b \\
&= K \cdot ((B \cdot D \cdot I) \cdot !a) \cdot !b \\
&= (B \cdot D \cdot I) \cdot !a \\
&= D \cdot (I \cdot !a) \\
&= D \cdot !a \\
&= a
\end{aligned}$$

□

3.4 Coproducts in restriction categories

3.4.1 Coproducts

Restriction categories may also have coproducts and initial objects.

Definition 3.4.1. In a restriction category \mathbb{X} , a coproduct is a *restriction coproduct* when the embeddings Π_0 and Π_1 are total.

Lemma 3.4.2. *The definition of restriction coproduct implies the following:*

- (i) $\overline{f + g} = \overline{f} + \overline{g}$ which means $+$ is a restriction functor.
- (ii) $\nabla : A + A \rightarrow A$ is total.
- (iii) $? : 0 \rightarrow A$ is total, where 0 is the initial object in the restriction category.

Proof.

(i) $+$ is a restriction functor. Consider the diagram:

$$\begin{array}{ccccc}
 A & \xrightarrow{f} & A' & & \\
 \searrow \Pi_0 & & \searrow \Pi'_0 & & \\
 & A + B & \xrightarrow{f+g} & A' + B' & \\
 \nearrow \Pi_1 & & \nearrow \Pi'_1 & & \\
 B & \xrightarrow{g} & B' & &
 \end{array}$$

In order to show $\overline{f+g} = \overline{f} + \overline{g}$, it suffices to show that $\Pi_0 \overline{f+g} = \Pi_0(\overline{f} + \overline{g}) = \overline{f} \Pi_0$.

$$\begin{aligned}
 \Pi_0 \overline{f+g} &= \overline{\Pi_0(f+g)} \Pi_0 && [\mathbf{R.4}] \\
 &= \overline{f \Pi'_0} \Pi_0 && \text{coproduct diagram} \\
 &= \overline{f \Pi'_0} \Pi_0 && \text{Lemma 2.2.3}[(iii)] \\
 &= \overline{f} \Pi_0 && \Pi'_0 \text{ total}
 \end{aligned}$$

(ii) $\nabla : A + A \rightarrow A$ is **total**. By the definition of $\nabla (= \langle 1|1 \rangle)$ and the co-product, the following diagram commutes,

$$\begin{array}{ccc}
 & A + A & \\
 \nearrow \Pi_0 & \downarrow \nabla & \nwarrow \Pi_1 \\
 A & = & A = A
 \end{array}$$

resulting in:

$$\begin{aligned}
 \Pi_0 \overline{\nabla} &= \overline{\Pi_0 \nabla} \Pi_0 \\
 &= \overline{1} \Pi_0 \\
 &= \Pi_0
 \end{aligned}$$

Similarly, $\Pi_1 \overline{\nabla} = \Pi_1$, hence, the restriction of ∇ is 1 and therefore ∇ is total.

(iii) $? : 0 \rightarrow A$ is **total**. This follows from

$$\begin{array}{ccc} 0 & \xrightarrow{\Pi_1} & A + 0 \\ & \searrow ? & \parallel \\ & & A \end{array}$$

so $?$ can be defined as the total coproduct injection.

□

Recall that when an object is both initial and terminal, it is referred to as a zero object and denoted as 0 . This gives rise to the zero map $0_{A,B} : A \rightarrow 0 \rightarrow B$ between any two objects.

Definition 3.4.3. Given a restriction category \mathbb{X} with a zero object, then 0 is a *restriction zero* when for each object A in \mathbb{X} we have $\overline{0_{A,A}} = 0_{A,A}$.

Lemma 3.4.4 (Cockett-Lack). *For a restriction category \mathbb{X} , the following are equivalent:*

- (i) \mathbb{X} has a restriction zero;
- (ii) \mathbb{X} has an initial object 0 and terminal object 1 and each initial map z_A is a restriction monic;
- (iii) \mathbb{X} has a terminal object 1 and each terminal map t_A is a restriction retraction.

3.4.2 Inverse categories with restriction coproducts

Proposition 3.4.5. *An inverse category \mathbb{X} with restriction coproducts is a preorder.*

Proof. By Lemma 3.4.2, we know ∇ is total and therefore $\nabla \nabla^{(-1)} = 1$. From the coproduct diagrams, we have $\Pi_0 \nabla = 1$ and $\Pi_1 \nabla = 1$. But this gives us $\nabla^{(-1)} \Pi_0^{(-1)} = \Pi_0 \nabla^{(-1)} = 1$ and $\nabla^{(-1)} \Pi_1^{(-1)} = 1$. Hence, $\nabla^{(-1)} = \Pi_0$ and $\nabla^{(-1)} = \Pi_1$.

This means for parallel maps $f, g : A \rightarrow B$, we have

$$f = \Pi_0 \langle f|g \rangle = \nabla^{(-1)} \langle f|g \rangle = \Pi_1 \langle f|g \rangle = g$$

and therefore \mathbb{X} is a preorder.

□

3.5 Disjointness in an inverse category

In the following, we will add two related structures to an inverse category with a restriction zero. This structure is meant to be evocative of the concept of *join* in a restriction category.

3.5.1 Disjointness relations

In this subsection, we will define a disjointness relationship between maps and explore alternate characterizations of this relation on the restriction idempotents of objects.

Definition 3.5.1. In an inverse category \mathbb{X} with a restriction zero, the relation \perp between two parallel maps $f, g : A \rightarrow B$ is called a *disjointness relation* when it satisfies the following properties:

- [Dis.1] For all $f : A \rightarrow B$, $f \perp 0$;
- [Dis.2] $f \perp g$ implies $\bar{f}g = 0$;
- [Dis.3] $f \perp g$, $f' \leq f$, $g' \leq g$ implies $f' \perp g'$;
- [Dis.4] $f \perp g$ implies $g \perp f$;
- [Dis.5] $f \perp g$ implies $hf \perp hg$; (Stable)
- [Dis.6] $f \perp g$ implies $\bar{f} \perp \bar{g}$ and $\hat{f} \perp \hat{g}$.
- [Dis.7] $\bar{f} \perp \bar{g}$, $\hat{h} \perp \hat{k}$ implies $fh \perp gk$;

Lemma 3.5.2. In Definition 3.5.1, provided we retain [Dis.1-5] we may replace [Dis.6] and [Dis.7] by:

$$[\mathbf{Dis.6}'] \quad f \perp g \text{ if and only if } \bar{f} \perp \bar{g} \text{ and } \hat{f} \perp \hat{g}.$$

Proof. Given [Dis.6] and [Dis.7], the *only if* direction of [Dis.6'] is immediate. To show the *if* direction, assume $\bar{f} \perp \bar{g}$ and $\hat{f} \perp \hat{g}$. This also means that $\bar{\bar{f}} \perp \bar{\bar{g}}$. Then, by [Dis.7], $\bar{\bar{f}}f \perp \bar{\bar{g}}g$ and therefore $f \perp g$.

Conversely, assume we are given **[Dis.6']**. Then, **[Dis.6]** follows immediately. To show **[Dis.7]**, assume we have $\bar{f} \perp \bar{g}$, $\hat{h} \perp \hat{k}$. As $\overline{fh} \leq \bar{f}$ and $\overline{gk} \leq \bar{g}$, by **[Dis.3]**, we know that $\overline{fh} \perp \overline{gk}$. Similarly, $\widehat{fh} \leq \hat{h}$ and $\widehat{gk} \leq \hat{k}$, giving us $\widehat{fh} \perp \widehat{gk}$. Then, from **[Dis.6']** we may conclude $fh \perp gk$, showing **[Dis.7]** holds. \square

Lemma 3.5.3. *In an inverse category \mathbb{X} with \perp a disjointness relation:*

- (i) $f \perp g$ if and only if $f^{(-1)} \perp g^{(-1)}$;
- (ii) $f \perp g$ implies $fh \perp gh$ (Universal);
- (iii) $f \perp g$ implies $f\hat{g} = 0$;
- (iv) if m, n are monic, then $fm \perp gn$ implies $\bar{f} \perp \bar{g}$;
- (v) if m, n are monic, then $m^{(-1)}f \perp n^{(-1)}g$ implies $\hat{f} \perp \hat{g}$;

Proof.

- (i) Assume $f \perp g$. Then we know that $\bar{f} \perp \bar{g}$ and $\hat{f} \perp \hat{g}$. But since $\hat{f} = \overline{f^{(-1)}}$ and $\bar{f} = \widehat{f^{(-1)}}$, this means $\overline{f^{(-1)}} \perp \overline{g^{(-1)}}$ and $\widehat{f^{(-1)}} \perp \widehat{g^{(-1)}}$ and again by the first item of this lemma, we have $f^{(-1)} \perp g^{(-1)}$. The converse follows with a similar argument.
- (ii) Assume $f \perp g$. By the previous item, we have $f^{(-1)} \perp g^{(-1)}$. By **[Dis.5]**, $h^{(-1)}f^{(-1)} \perp h^{(-1)}g^{(-1)}$, giving us $(fh)^{(-1)} \perp (gh)^{(-1)}$. Again by the previous item, we now have $fh \perp gh$.
- (iii) Assume $f \perp g$. From item (i) and reflexivity, we know that $g^{(-1)} \perp f^{(-1)}$ and therefore $\overline{g^{(-1)}}f^{(-1)} = \hat{g}f^{(-1)} = 0$. However, in an inverse category, $0^{(-1)} = 0$ and therefore $0 = (\hat{g}f^{(-1)})^{(-1)} = f\hat{g}^{(-1)} = f\hat{g}$.
- (iv) Assume we have $fm \perp gn$ where m, n are monic. By **[Dis.6]**, $\overline{fm} \perp \overline{gn}$. By Lemma 2.2.3, $\overline{fm} = \overline{f\overline{m}} = \overline{f}1 = \bar{f}$ and therefore $\bar{f} \perp \bar{g}$.

(v) This is a corollary to the previous item. By assumption, we have $m^{(-1)}f \perp n^{(-1)}g$ and therefore $f^{(-1)}m \perp g^{(-1)}n$. By the previous item, this means $\overline{f^{(-1)}} \perp \overline{g^{(-1)}}$ and hence $\hat{f} \perp \hat{g}$.

□

We may define the disjointness relation via its action in $\mathcal{O}(a)$.

Definition 3.5.4. Given an inverse category \mathbb{X} , a relation $\perp_A \subseteq \mathcal{O}(A)^2$ for each $A \in \text{ob}(\mathbb{X})$, is an *open disjointness* relation when for all $e, e' \in \mathcal{O}(A)$

$$[\mathcal{O}\text{dis.1}] \quad 1 \perp_A 0;$$

$$[\mathcal{O}\text{dis.2}] \quad e \perp_A e' \text{ implies } e' \perp_A e;$$

$$[\mathcal{O}\text{dis.3}] \quad e \perp_A e' \text{ implies } ee' = 0;$$

$$[\mathcal{O}\text{dis.4}] \quad e \perp_A e' \text{ implies } \overline{fe} \perp_B \overline{fe'} \text{ for all } f : B \rightarrow A;$$

$$[\mathcal{O}\text{dis.5}] \quad e \perp_A e' \text{ implies } \widehat{eg} \perp_C \widehat{e'g} \text{ for all } g : A \rightarrow C;$$

$$[\mathcal{O}\text{dis.6}] \quad e \perp_A e', e_1 \leq e, e'_1 \leq e' \text{ implies } e_1 \perp_A e'_1.$$

We will normally write \perp rather than \perp_A where the object is either clear or not germane to the point under discussion.

Proposition 3.5.5. *If \perp is a disjointness relation in \mathbb{X} , it is an open disjointness relation on the restriction idempotents.*

Proof.

[$\mathcal{O}\text{dis.1}$] This follows immediately from [**Dis.1**] by taking $f = 1$.

[$\mathcal{O}\text{dis.2}$] Reflexivity follows directly from [**Dis.4**].

[$\mathcal{O}\text{dis.3}$] By [**Dis.2**], $0 = \bar{e}e' = ee'$.

[$\mathcal{O}\text{dis.4}$] Given $e \perp e'$, we have $fe \perp fe'$ by [**Dis.5**]. Then, by [**Dis.6**] we may conclude $\overline{fe} \perp \overline{fe'}$.

[**Odis.5**] This follows from the above item, using $g^{(-1)}$ for f . This means we have $\overline{g^{(-1)}e} \perp \overline{g^{(-1)}e'}$. But this gives us $\overline{(eg)^{(-1)}} \perp \overline{(e'g)^{(-1)}}$. Recalling from Lemma 2.2.11 that $\hat{k} = \overline{k^{(-1)}}$, we may conclude $\widehat{eg} \perp \widehat{e'g}$.

[**Odis.6**] Assuming $e \perp e'$ and $e_1 \leq e$, $e'_1 \leq e'$, by [**Dis.3**], $e_1 \perp e'_1$.

Therefore, \perp acts as an open disjointness relation on $\mathcal{O}(A)^2$.

□

Definition 3.5.6. If $\underline{\perp}$ is an open disjointness relation in \mathbb{X} , then we may define a relation

${}_A\perp_B \in \mathbb{X}(A, B)^2$ by

$$\frac{f, g : A \rightarrow B, \bar{f} \underline{\perp} \bar{g}, \hat{f} \underline{\perp} \hat{g}}{f {}_A\perp_B g}.$$

We call ${}_A\perp_B$ an *extended disjointness relation*.

Proposition 3.5.7. If $\underline{\perp}$ is an extended disjointness relation based on $\underline{\perp}$ in \mathbb{X} , then \perp is a disjointness relation in \mathbb{X} .

Proof.

[**Dis.1**] We need to show $f \perp 0$ for any f . We know that $1 \underline{\perp} 0$ and therefore $\bar{f} \underline{\perp} 0$ and $\hat{f} \underline{\perp} 0$, as $\bar{f} \leq 1$ and $\hat{f} \leq 1$. This gives us $f \perp 0$.

[**Dis.2**] Assume $f \perp g$, i.e., $\bar{f} \underline{\perp} \bar{g}$. Then, $\bar{f}g = \bar{f}\bar{g}g = 0g = 0$.

[**Dis.3**] We are given $f \perp g$, $f' \leq f$ and $g' \leq g$. By lemma 2.2.9[(iii)] $\bar{f}' \leq \bar{f}$ and $\bar{g}' \leq \bar{g}$. Then, by [**Odis.6**], as $\bar{f} \underline{\perp} \bar{g}$ we have $\bar{f}' \underline{\perp} \bar{g}'$. By 2.2.9[(iv)], we have $\hat{f}' \leq \hat{f}$ and $\hat{g}' \leq \hat{g}$. Then, by [**Odis.6**], as $\hat{f} \underline{\perp} \hat{g}$ we have $\hat{f}' \underline{\perp} \hat{g}'$. This means $f' \perp g'$.

[**Dis.4**] Reflexivity of \perp follows immediately from the reflexivity of $\underline{\perp}$.

[**Dis.5**] Assume $f \perp g$, i.e., $\bar{f} \underline{\perp} \bar{g}$ and $\hat{f} \underline{\perp} \hat{g}$. Then we have $\overline{hf} \underline{\perp} \overline{hg}$ by [**Odis.4**]. By lemma 2.2.9[(ii)] we have $\widehat{hf} \leq \hat{f}$ and $\widehat{hg} \leq \hat{g}$. Therefore we have $\widehat{hf} \underline{\perp} \widehat{hg}$ by [**Odis.6**] and therefore $hf \perp hg$.

[Dis.6] This follows directly from definition 3.5.6.

[Dis.7] We assume $\bar{f} \perp \bar{g}$ and $\hat{h} \perp \hat{k}$. By definition 3.5.6 we have $\bar{f} \perp\!\!\!\perp \bar{g}$ and $\hat{h} \perp\!\!\!\perp \hat{k}$. By lemma 2.2.9(i), we have $\overline{fh} \leq \bar{f}$ and $\overline{gk} \leq \bar{g}$. Therefore, $\overline{fh} \perp\!\!\!\perp \overline{gk}$ by [Odis.6]. By 2.2.9(ii), $\widehat{fh} \leq \hat{h}$ and $\widehat{gk} \leq \hat{k}$, giving us $\widehat{fh} \perp\!\!\!\perp \widehat{gk}$ also by [Odis.6]. This means $fh \perp gk$.

□

Theorem 3.5.8. *To give a disjointness relation \perp on \mathbb{X} is to give an open disjointness relation $\perp\!\!\!\perp$ on \mathbb{X} .*

Proof. Suppose we are given the disjointness relation \perp . By Proposition 3.5.5, this is an open disjointness relation on each of the sets of idempotents, $\mathcal{O}(A)$. We will label that relation $\perp\!\!\!\perp$.

Use Definition 3.5.6 to create an extended disjointness relation based on $\perp\!\!\!\perp$, signify it by $\underline{\perp\!\!\!\perp}$. By Proposition 3.5.7, $\underline{\perp\!\!\!\perp}$ is a disjointness relation on \mathbb{X} .

Assume $f \perp g$. Then we have $\bar{f} \perp\!\!\!\perp \bar{g}$ and $\hat{f} \perp\!\!\!\perp \hat{g}$ by [Dis.6] and Proposition 3.5.5. Then, from Definition 3.5.6, we have $f \underline{\perp\!\!\!\perp} g$.

Assume $f \underline{\perp\!\!\!\perp} g$. Then we must have had $\bar{f} \perp\!\!\!\perp \bar{g}$ and $\hat{f} \perp\!\!\!\perp \hat{g}$ by Definition 3.5.6 and therefore $\bar{f} \perp \bar{g}$ and $\hat{f} \perp \hat{g}$. Then, by Proposition 3.5.3, we have $f \perp g$.

Now, suppose we are given the open disjointness relation $\perp\!\!\!\perp$. Similar to above, we can construct the extended disjointness relation \perp by Definition 3.5.6. From the disjointness relation \perp , we have the open disjointness relation $\overline{\perp}$ by Lemma 3.5.5.

Assume $e \underline{\perp\!\!\!\perp} e'$. As this means both $\bar{e} \perp\!\!\!\perp \bar{e}'$ and $\hat{e} \perp\!\!\!\perp \hat{e}'$, we have $e \perp e'$. By Proposition 3.5.5 this means $e \overline{\perp} e'$.

If we are given that $e \overline{\perp} e'$, then we know that $e \perp e'$ by Proposition 3.5.5. From Definition 3.5.6, this requires that $\bar{e} \perp\!\!\!\perp \bar{e}'$ and $\hat{e} \perp\!\!\!\perp \hat{e}'$, but that just means $e \underline{\perp\!\!\!\perp} e'$. □

Note that while we have worked with binary disjointness throughout this section, one

may extend the concept to lists of maps simply by considering disjointness pairwise. I.e., we have $\perp [f_1, f_2, \dots, f_n]$ if and only if $f_i \perp f_k$ whenever $i \neq j$.

Disjointness is additional structure on a restriction category, i.e., it is possible to have more than one disjointness relation on the category.

Example 3.5.9. Consider the restriction category **INJ**. Here, the objects are sets and maps are the partial injective set functions, where $\bar{f} = id_{|\text{dom}(f)}$. The restriction zero is the empty map (i.e., $\text{dom}(0) = \text{range}(0) = \emptyset$).

We may define the disjointness relation \perp by $f \perp g$ if and only if $\text{dom}(f) \cap \text{dom}(g) = \emptyset$ and $\text{range}(f) \cap \text{range}(g) = \emptyset$. It is reasonably straightforward to verify **[Dis.1]** through **[Dis.7]**. For example, take **[Dis.7]**:

Proof. We are given $\bar{f} \perp \bar{g}$ and $\hat{h} \perp \hat{k}$. This means

$$\text{dom } f \cap \text{dom } g = \emptyset \text{ and } \text{range } h \cap \text{range } k = \emptyset.$$

Note that in general for partial injective functions m and n we have $\text{dom } mn \subseteq \text{dom } m$ and that $\text{range } mn \subseteq \text{range } n$. Hence we have

$$\begin{aligned} \text{dom } fh \cap \text{dom } gk &\subseteq \text{dom } f \cap \text{dom } g = \emptyset \\ \text{range } fh \cap \text{range } gk &\subseteq \text{range } h \cap \text{range } k = \emptyset. \end{aligned}$$

Therefore, $fh \perp gk$. □

We may define a different disjointness relation, \perp' , on the same restriction category. Define $f \perp' g$ if and only if one of f or g is the restriction 0, \emptyset . As $0 = \bar{0} = \hat{0} = h0 = 0k$, all of the seven disjointness axioms are easily verifiable.

Although disjointness is additional structure on a restriction category, one can use the disjointness structure of a base category (or categories) to define a disjointness structure on derived categories, such as the product category.

Lemma 3.5.10. *If \mathbb{X} and \mathbb{Y} are inverse categories with restriction zeros and respective disjointness relations \perp and \perp' , then we may construct a disjointness relation \perp_{\times} on $\mathbb{X} \times \mathbb{Y}$.*

Proof. Recall that product categories are defined component-wise. These definitions extend to the restriction, the inverse and the restriction zero. That is:

- If (f, g) is a map in $\mathbb{X} \times \mathbb{Y}$, then $(f, g)^{(-1)} = (f^{(-1)}, g^{(-1)})$;
- If (f, g) is a map in $\mathbb{X} \times \mathbb{Y}$, then $\overline{(f, g)} = (\overline{f}, \overline{g})$;
- The map $(0_X, 0_Y)$ is the restriction zero in $\mathbb{X} \times \mathbb{Y}$.

Following this pattern, for (f, g) and (h, k) maps in $\mathbb{X} \times \mathbb{Y}$, $(f, g) \perp_{\times} (h, k)$ iff $f \perp h$ and $g \perp' k$.

Verifying the disjointness axioms is straightforward, we show axioms 2 and 5. Proofs of the others are similar.

[Dis.2] : Given $(f, g) \perp_{\times} (h, k)$, we have $\overline{(f, g)}(h, k) = (\overline{f}, \overline{g})(h, k) = (\overline{f}h, \overline{g}k) = (0, 0) = 0$.

[Dis.5] : We are given $(f, g) \perp_{\times} (h, k)$. Consider the map $z = (x, y)$ in $\mathbb{X} \times \mathbb{Y}$. We know that $xf \perp xh$ and $yg \perp yk$, therefore we have $z(f, g) = (xf, yg) \perp_{\times} (xh, yk) = z(h, k)$.

□

3.5.2 Disjoint joins

We now consider additional structure on the inverse category, dependant upon the disjointness relation.

Definition 3.5.11. An *inverse category with disjoint joins* is an inverse category \mathbb{X} , with a restriction 0, a disjointness relation \perp and a binary operator on disjointness parallel maps:

$$\frac{f : A \rightarrow B, \quad g : A \rightarrow B, \quad f \perp g}{f \sqcup g : A \rightarrow B}$$

where the following hold:

$$[\mathbf{DJ.1}] \quad f \leq f \sqcup g \text{ and } g \leq f \sqcup g;$$

$$[\mathbf{DJ.2}] \quad f \leq h, g \leq h \text{ and } f \perp g \text{ implies } f \sqcup g \leq h;$$

$$[\mathbf{DJ.3}] \quad h(f \sqcup g) = hf \sqcup hg. \text{ (Stable)}$$

$$[\mathbf{DJ.4}] \quad \perp [f, g, h] \text{ if and only if } f \perp (g \sqcup h).$$

The binary operator, \sqcup , is referred to as the *disjoint join*.

Note that $[\mathbf{DJ.1}]$ with $[\mathbf{DJ.2}]$ immediately gives us that there is only one disjoint join given a specific disjointness relation.

Lemma 3.5.12. *Suppose \mathbb{X} in an inverse category with disjoint joins, with the join \sqcup and that it has a second disjoint join, \square . Then $f \sqcup g = f \square g$ for all maps f, g in \mathbb{X} .*

Proof. The first axiom tells us:

$$f, g \leq f \sqcup g \text{ and } f, g \leq f \square g.$$

Using the second axiom, we may therefore conclude $f \sqcup g \leq f \square g$ and $f \square g \leq f \sqcup g$, hence $f \sqcup g = f \square g$. \square

Lemma 3.5.13. *In an inverse category with disjoint joins, the disjoint join respects the restriction and is universal. Additionally, it is a partial associative and commutative operation, with identity 0. That is, the following hold:*

$$(i) \quad \overline{f \sqcup g} = \overline{f} \sqcup \overline{g};$$

$$(ii) \quad (f \sqcup g)k = fk \sqcup gk \text{ (Universal);}$$

$$(iii) \quad f \perp g, g \perp h, f \perp h \text{ implies that } (f \sqcup g) \sqcup h = f \sqcup (g \sqcup h);$$

$$(iv) \quad f \perp g \text{ implies } f \sqcup g = g \sqcup f;$$

$$(v) \ f \sqcup 0 = f.$$

Proof.

- (i) As $\bar{f}, \bar{g} \leq \overline{f \sqcup g}$, we immediately have $\bar{f} \sqcup \bar{g} \leq \overline{f \sqcup g}$. To show the other direction, consider

$$\begin{aligned} \bar{f}(\bar{f} \sqcup \bar{g})(f \sqcup g) &= (\bar{f} \bar{f} \sqcup \bar{f} \bar{g})(f \sqcup g) && [\mathbf{DJ.3}] \\ &= \bar{f}(f \sqcup g) && \text{Lemma 2.2.3, } [\mathbf{Dis.2}] \\ &= f. \end{aligned}$$

Hence, we have $f \leq (\bar{f} \sqcup \bar{g})(f \sqcup g)$ and similarly, so is g . By $[\mathbf{DJ.2}]$ and that $\bar{f} \sqcup \bar{g}$ is a restriction idempotent, we then have

$$f \sqcup g \leq (\bar{f} \sqcup \bar{g})(f \sqcup g) \leq f \sqcup g$$

and therefore $f \sqcup g = (\bar{f} \sqcup \bar{g})(f \sqcup g)$. By Lemma 2.2.4, $\overline{f \sqcup g} \leq \bar{f} \sqcup \bar{g}$ and so $\overline{f \sqcup g} = \bar{f} \sqcup \bar{g}$.

- (ii) First consider when f, g and k are restriction idempotents, say e_0, e_1 and e_2 . Then, we have $(e_0 \sqcup e_1)e_2 = e_2(e_0 \sqcup e_1) = e_2e_0 \sqcup e_2e_1 = e_0e_2 \sqcup e_1e_2$. Next, note that for general f, g, h , we have $fk \sqcup gk \leq (f \sqcup g)k$ as both $fk, gk \leq (f \sqcup g)k$. By Lemma 2.2.4, we need only show that their restrictions are equal:

$$\begin{aligned} \overline{(f \sqcup g)k} &= \overline{\bar{f} \sqcup \bar{g}(f \sqcup g)k} && [\mathbf{R.1}] \\ &= \overline{\bar{f} \sqcup \bar{g}} \overline{(f \sqcup g)k} && [\mathbf{R.3}] \\ &= (\bar{f} \sqcup \bar{g}) \overline{(f \sqcup g)k} && \text{previous item} \\ &= \bar{f} \overline{(f \sqcup g)k} \sqcup \bar{g} \overline{(f \sqcup g)k} && \text{idempotent universal} \\ &= \overline{\bar{f}(f \sqcup g)k} \sqcup \overline{\bar{g}(f \sqcup g)k} && [\mathbf{R.3}] \\ &= \overline{fk} \sqcup \overline{gk} \\ &= \overline{fk \sqcup gk}. \end{aligned}$$

Therefore, as the restrictions are equal, we have shown $(f \sqcup g)k = fk \sqcup gk$.

(iii) *Associativity*: Note that [DJ.4] shows that both sides of the equation exist.

To show they are equal, we show that they are less than or equal to each other.

From the definitions, we know that $f \sqcup g, h \leq (f \sqcup g) \sqcup h$, which also means

$f, g \leq (f \sqcup g) \sqcup h$. Similarly, $g \sqcup h \leq (f \sqcup g) \sqcup h$ and then $f \sqcup (g \sqcup h) \leq (f \sqcup g) \sqcup h$.

Conversely, $f, g, h \leq f \sqcup (g \sqcup h)$ and therefore $(f \sqcup g) \sqcup h \leq f \sqcup (g \sqcup h)$ and both sides are equal.

(iv) *Commutativity*: Note first that both f and g are less than or equal to both

$f \sqcup g$ and $g \sqcup f$, by [DJ.1]. By [DJ.2], we have $f \sqcup g \leq g \sqcup f$ and $g \sqcup f \leq f \sqcup g$

and we may conclude $f \sqcup g = g \sqcup f$.

(v) *Identity*: By [DJ.1], $f \leq f \sqcup 0$. As $0 \leq f$ and $f \leq f$, by [DJ.2], $f \sqcup 0 \leq f$

and we have $f = f \sqcup 0$.

□

Note that the previous lemma and proof of associativity allows a simple inductive argument which shows that having binary disjoint joins extends unambiguously to disjoint joins of an arbitrary finite collection of disjoint maps.

We will write $[f_i]$ to signify a list of maps, where each $f_i : A \rightarrow B$. For disjointness, $\perp [f_i]$ will mean that $f_j \perp f_k$ where $j \neq k$ and $f_j, f_k \in [f_i]$. Finally, $\sqcup [f_i]$ will mean the disjoint join of all maps f_i , i.e., $f_1 \sqcup f_2 \sqcup \cdots \sqcup f_n$.

Lemma 3.5.14. *In an inverse category with disjoint joins, $\perp [f_i]$ if and only if $\sqcup [f_i]$ is defined unambiguously.*

Proof. Using [Dj.4], proceed as in the proof of Lemma 3.5.13[(iii)], inducting on n . □

Lemma 3.5.15. *Given \mathbb{X} is an inverse category with a disjoint join, then if $f_i, g_j : A \rightarrow B$ and $\perp [f_i]$ and $\perp [g_j]$, then $\sqcup [f_i] \perp \sqcup [g_j]$ if and only if $f_i \perp g_j$ for all i, j ;*

Proof. Assume $\sqcup[f_i] \perp \sqcup[g_j]$. Then by [Dj.4] and associativity, we have $\sqcup[f_i] \perp g_j$ for each j . Then, applying the reflexivity of \perp , [Dj.4] and associativity, we have $f_i \perp g_j$ for each i and j .

Assume $f_i \perp g_j$ for each i and j . Then by [Dj.4] and associativity, $f_i \perp \sqcup[g_j]$ for each i . Applying [Dj.4] again, we have $\sqcup[f_i] \perp \sqcup[g_j]$. \square

Following the same method as in the previous section, we show that the product of two inverse categories with disjoint joins has a disjoint join.

Lemma 3.5.16. *Given \mathbb{X}, \mathbb{Y} are inverse categories with disjoint joins, \sqcup and \sqcup' respectively, then the category $\mathbb{X} \times \mathbb{Y}$ is an inverse category with disjoint joins.*

Proof. From Lemma 3.5.10, we know $\mathbb{X} \times \mathbb{Y}$ has a disjointness relation that is defined point-wise. We therefore define \sqcup_{\times} the disjoint join on $\mathbb{X} \times \mathbb{Y}$ by

$$(f, g) \sqcup_{\times} (h, k) = (f \sqcup h, g \sqcup' k) \quad (3.12)$$

We now prove each of the axioms in Definition 3.5.11 hold.

[Dj.1] From Equation (equation (3.12)), we see that since $f, h \leq f \sqcup h$ and $g, k \leq g \sqcup' k$, we have $(f, g) \leq (f, g) \sqcup_{\times} (h, k)$ and $(h, k) \leq (f, g) \sqcup_{\times} (h, k)$.

[Dj.2] Suppose $(f, g) \leq (x, y)$, $(h, k) \leq (x, y)$ and $(f, g) \perp_{\times} (h, k)$. Then regarding it point-wise, we have $(f, g) \sqcup_{\times} (h, k) = (f \sqcup h, g \sqcup' k) \leq (x, y)$.

[Dj.3] $(x, y) ((f, g) \sqcup_{\times} (h, k)) = (x(f \sqcup h), y(g \sqcup' k)) = (xf \sqcup xh, yg \sqcup' yk) = (xf, yg) \sqcup_{\times} (xh, yk) = ((x, y)(f, g)) \sqcup_{\times} ((x, y)(h, k))$.

[Dj.4] Given $\perp_{\times} [(f, g), (h, k), (x, y)]$, we know $f \perp (h \sqcup x)$ and $g \perp' (k \sqcup' y)$. Hence, $(f, g) \perp_{\times} ((h, k) \sqcup_{\times} (x, y))$. The opposite direction is similar.

\square

3.5.3 Monoidal Tensors for disjointness

Suppose we are given a monoidal tensor \oplus on \mathbb{X} , an inverse category with a restriction zero. Under certain conditions, it is possible to define disjointness based upon the action of the tensor. Note that throughout, we are assuming the following naming for the standard monoidal tensor isomorphisms.

$$\begin{aligned} u_{\oplus}^l &: 0 \oplus A \rightarrow A \\ u_{\oplus}^r &: A \oplus 0 \rightarrow A \\ a_{\oplus} &: (A \oplus B) \oplus C \rightarrow A \oplus (B \oplus C) \\ c_{\oplus} &: A \oplus B \rightarrow B \oplus A. \end{aligned}$$

Note we also require the tensor isomorphisms above be natural.

Definition 3.5.17. Suppose we are given an inverse category \mathbb{X} with restriction zero and a symmetric monoidal tensor \oplus . Then \oplus is a *disjointness tensor* when:

- It is a restriction functor — i.e., $_{\oplus} : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{X}$.
- The unit is the restriction zero. ($0 : \mathbf{1} \rightarrow \mathbb{X}$ picks out the restriction zero in \mathbb{X}).
- Define $\Pi_0 := (1 \oplus 0)u_{\oplus}^r : A \oplus B \rightarrow A$ and $\Pi_1 := (0 \oplus 1)u_{\oplus}^l : A \oplus B \rightarrow B$. Then Π_0 and Π_1 are jointly monic. That is, whenever $f\Pi_0 = g\Pi_0$ and $f\Pi_1 = g\Pi_1$ then $f = g$.
- Define $\Pi_0 = u_{\oplus}^{r(-1)}(1 \oplus 0) : A \rightarrow A \oplus A$ and $\Pi_1 = u_{\oplus}^{l(-1)}(0 \oplus 1) : A \rightarrow A \oplus A$. Then Π_0 and Π_1 are jointly epic. That is, if $\Pi_0 f = \Pi_0 g$ and $\Pi_1 f = \Pi_1 g$, then $f = g$.

Lemma 3.5.18. Given an inverse category \mathbb{X} with restriction zero and disjointness tensor \oplus , then the map $0 \oplus 0 : A \oplus B \rightarrow C \oplus D$ is the map $0 : A \oplus B \rightarrow C \oplus D$.

Proof. Recall the zero map factors through the restriction zero, i.e. $0 : A \rightarrow B$ is the same as saying $A \xrightarrow{!} 0 \xrightarrow{?} B$. Additionally, as objects, $0 \oplus 0 \cong 0$ — the restriction zero.

Therefore the map $0 \oplus 0 : A \oplus B \rightarrow C \oplus D$ is writable as

$$A \oplus B \xrightarrow{! \oplus !} 0 \oplus 0 \xrightarrow{? \oplus ?} C \oplus D,$$

which may then be rewritten as

$$A \oplus B \xrightarrow{! \oplus !} 0 \oplus 0 \xrightarrow{u_{\oplus}^l} 0 \xrightarrow{u_{\oplus}^{l(-1)}} 0 \oplus 0 \xrightarrow{? \oplus ?} C \oplus D.$$

But by the properties of the restriction zero, $(! \oplus !)u_{\oplus}^l = !$ and $u_{\oplus}^{l(-1)}(? \oplus ?) = ?$ and therefore the map $0 \oplus 0 : A \oplus B \rightarrow C \oplus D$ is the same as the map $0 : A \oplus B \rightarrow C \oplus D$. \square

Lemma 3.5.19. *Given an inverse category \mathbb{X} with a restriction zero and a disjointness tensor, the map Π_0 is natural in the left component and Π_1 is natural in the right, up to isomorphism. This means:*

$$\Pi_0(f \oplus g) = f \Pi_0 \quad \text{and} \quad \Pi_1(f \oplus g) = g \Pi_1.$$

Proof. For the left and right naturality, we see:

$$\Pi_0(f \oplus g) = u_{\oplus}^{r(-1)}(1 \oplus 0)(f \oplus g) = u_{\oplus}^{r(-1)}(f \oplus 0) = f u_{\oplus}^{r(-1)}(1 \oplus 0) = f \Pi_0,$$

and

$$\Pi_1(f \oplus g) = u_{\oplus}^{l(-1)}(0 \oplus 1)(f \oplus g) = u_{\oplus}^{l(-1)}(0 \oplus g) = g u_{\oplus}^{l(-1)}(0 \oplus 1) = g \Pi_1.$$

\square

Lemma 3.5.20. *Given an inverse category \mathbb{X} with restriction zero and disjointness tensor \oplus , then the following hold:*

1. $\Pi_i \Pi_i = \overline{\Pi_i}$ and $\Pi_i \Pi_i = \overline{\Pi_i} = 1$;
2. $\overline{\Pi_0} \Pi_1 = 0$ and $\overline{\Pi_1} \Pi_0 = 0$;

3. $\Pi_1\Pi_0 = 0$, $\Pi_1\overline{\Pi_0} = 0$, $\Pi_0\Pi_1 = 0$ and $\Pi_0\overline{\Pi_1} = 0$;

4. the maps Π_0 and Π_1 are monic.

Proof. For item 1, recalling that the restriction zero is its own partial inverse, we see from their definitions that $\Pi_i = \Pi_i^{(-1)}$. Calculating the restriction of Π_0 ,

$$\Pi_0\Pi_0 = u_{\oplus}^r{}^{(-1)}(1 \oplus 0)(1 \oplus 0)u_{\oplus}^r = (u_{\oplus}^r{}^{(-1)}(1 \oplus 0))u_{\oplus}^r = 1u_{\oplus}^r{}^{(-1)}u_{\oplus}^r = 1.$$

The calculation for Π_1 and Π_1 are analogous.

For the second item,

$$\overline{\Pi_0}\Pi_1 = \overline{(1 \oplus 0)u_{\oplus}^r}(0 \oplus 1)u_{\oplus}^l = \overline{1 \oplus 0}(0 \oplus 1)u_{\oplus}^l = (1 \oplus 0)(0 \oplus 1)u_{\oplus}^l = (0 \oplus 0)u_{\oplus}^l = 0,$$

and

$$\overline{\Pi_1}\Pi_0 = \overline{(0 \oplus 1)u_{\oplus}^l}(1 \oplus 0)u_{\oplus}^r = (0 \oplus 1)(1 \oplus 0)u_{\oplus}^r = (0 \oplus 0)u_{\oplus}^r = 0.$$

For the third item, we see

$$\Pi_0\Pi_1 = (u_{\oplus}^r{}^{(-1)}(1 \oplus 0))(0 \oplus 1)u_{\oplus}^l = u_{\oplus}^r{}^{(-1)}(0 \oplus 0)u_{\oplus}^l = 0$$

and

$$\Pi_1\Pi_0 = (u_{\oplus}^l{}^{(-1)}(0 \oplus 1))(1 \oplus 0)u_{\oplus}^r = u_{\oplus}^l{}^{(-1)}(0 \oplus 0)u_{\oplus}^r = 0.$$

As $\overline{\Pi_0} = 1 \oplus 0$ and $\overline{\Pi_1} = 0 \oplus 1$, we see the other two identities hold as well.

To prove Π_0 is monic, suppose $f\Pi_0 = g\Pi_0$. Therefore we must have

$$f = f(\Pi_0\Pi_0) = (f\Pi_0)\Pi_0 = (g\Pi_0)\Pi_0 = g(\Pi_0\Pi_0) = g.$$

The proof that Π_1 is monic is similar. □

Corollary 3.5.21. *In an inverse category \mathbb{X} with a restriction zero and disjointness tensor, the following hold:*

- (i) $\Pi_0 f \oplus g \Pi_0 = f$; (iii) $\Pi_1 f \oplus g \Pi_0 = 0$;
(ii) $\Pi_0 f \oplus g \Pi_1 = 0$; (iv) $\Pi_1 f \oplus g \Pi_1 = g$.

Additionally, if t is a map such that for $i \in \{0, 1\}$,

$$\Pi_i t \Pi_j = \begin{cases} t_i & : i \neq j \\ 0 & : i = j, \end{cases}$$

then $t = t_0 \oplus t_1$

Proof. The calculations for $f \oplus g$ follow from Lemma 3.5.19 and Lemma 3.5.20. For example, $\Pi_0 f \oplus g \Pi_0 = f \Pi_0 \Pi_0 = f$.

For the second claim, note that we have $\Pi_0(t \Pi_0) = t_0 = \Pi_0(t_0 \oplus t_1) \Pi_0$ and $\Pi_1(t \Pi_0) = 0 = \Pi_1(t_0 \oplus t_1) \Pi_0$, hence $t \Pi_0 = (t_0 \oplus t_1) \Pi_0$. Similary, we see $t \Pi_1 = (t_0 \oplus t_1) \Pi_1$ and therefore $t = t_0 \oplus t + 1$. □

Definition 3.5.22. In an inverse category \mathbb{X} with a restriction zero and disjointness tensor, we define two partial operations on pairs of arrows in \mathbb{X} to another arrow in \mathbb{X} . First, for arrows $f : A \rightarrow B$ and $g : A \rightarrow C$, we define $f \nabla g$ as being the map that makes diagram (equation (3.13)) below commute, when it exists.

$$\begin{array}{ccccc} & & B & \xleftarrow{\Pi_0} & B \oplus C & \xrightarrow{\Pi_1} & C \\ & & \swarrow g & & \uparrow f \nabla g & & \searrow f \\ & & A & & & & \end{array} \quad (3.13)$$

Then for $h : B \rightarrow A$, $k : C \rightarrow A$, $h \triangle k$ is that map that makes diagram (equation (3.14)) commute, if it exists.

$$\begin{array}{ccccc} & & A & & \\ & \nearrow h & \uparrow h \triangle k & \nwarrow k & \\ B & \xrightarrow{\Pi_0} & B \oplus C & \xleftarrow{\Pi_1} & C \end{array} \quad (3.14)$$

Due to Π_0 and Π_1 being jointly monic, $f \nabla g$ is unique when it exists. Similarly, as Π_0 and Π_1 are jointly epic, $f \Delta g$ is unique when it exists.

We give a lemma exploring the behaviour of the two operations: ∇ and Δ .

Lemma 3.5.23. *Given \mathbb{X} is an inverse category with a restriction zero and a disjointness tensor \oplus then the following relations hold for ∇ and Δ :*

- (i) *If $f \nabla g$ exists, then $g \nabla f$ exists. If $f \Delta g$ exists, then $g \Delta f$ exists.*
- (ii) *$f \nabla 0$ and $f \Delta 0$ always exist.*
- (iii) *When $f \nabla g$ exists, $\bar{f}(f \nabla g) = f \nabla 0$, $\bar{f}g = 0$, $\bar{g}(f \nabla g) = 0 \nabla g$ and $\bar{g}f = 0$.*
- (iv) *Dually to the previous item, when $f \Delta g$ exists, $(f \Delta g)\hat{f} = f \Delta 0$, $g\hat{f} = 0$, $(f \Delta g)\hat{g} = 0 \Delta g$ and $f\hat{g} = 0$.*
- (v) *When $f \nabla g$ exists, $f \nabla g(h \oplus k) = fh \nabla gk$.*
- (vi) *Dually, when $f \Delta g$ exists, $(h \oplus k)f \Delta g = hf \Delta kg$.*
- (vii) *When $f \nabla g$ exists, then $h(f \nabla g) = hf \nabla hg$ and when $f \Delta g$ exists, $(f \Delta g)h = fh \Delta gh$.*
- (viii) *If $\bar{f} \nabla \bar{g}$ exists, then $\bar{f} \Delta \bar{g}$ exists and is the partial inverse of $\bar{f} \nabla \bar{g}$.*
- (ix) *If $f \nabla g$ exists and $f' \leq f$, $g' \leq g$, then $f' \nabla g'$ exists.*
- (x) *When $f \Delta g$ exists, $(f \Delta g)(f \Delta g)^{(-1)} = \bar{f} \oplus \bar{g}$.*
- (xi) *Given $f \nabla g$ and $h \nabla k$ exist, then $(f \oplus h) \nabla (g \oplus k) = (f \nabla g) \oplus (h \nabla k)$. Dually, the existence of $f \Delta g$ and $h \Delta k$ implies $(f \oplus h) \Delta (g \oplus k) = (f \Delta g) \oplus (h \Delta k)$.*

Proof.

- (i) $g \nabla f = (f \nabla g)c_{\oplus}$ and $g \Delta f = c_{\oplus}(f \Delta g)$.

(ii) Consider $f\Pi_0$. Then $f\Pi_0\Pi_0 = f$ and $f\Pi_0\Pi_1 = f0 = 0$. Hence, $f\Pi_0 = f \nabla 0$.
 Consider $\Pi_0 f$. Then $\Pi_0\Pi_0 f = f$ and $\Pi_1\Pi_0 f = 0f = 0$ and therefore $\Pi_0 f = (f \triangle 0)$.

(iii) Using Lemma 3.5.20

$$\overline{f}g = \overline{(f \nabla g)\Pi_0}(f \nabla g)\Pi_1 = (f \nabla g)\overline{\Pi_0}\Pi_1 = 0.$$

Similarly, $\overline{g}f = f \nabla g\overline{\Pi_1}\Pi_0 = 0$.

Recall that Π_0 and Π_1 are jointly monic. We have $\overline{f}(f \nabla g)\Pi_0 = \overline{f}f = f = (f \nabla 0)\Pi_0$ and $\overline{f}(f \nabla g)\Pi_1 = \overline{f}g = 0 = (f \nabla 0)\Pi_1$. Therefore, $\overline{f}(f \nabla g) = f \nabla 0$.
 Similarly, $\overline{g}(f \nabla g) = 0 \nabla g$.

(iv) Using Lemma 3.5.20

$$\begin{aligned} g\hat{f} &= \Pi_1(f \triangle g)(\Pi_0(\widehat{f \triangle g})) = \Pi_1(f \triangle g)\overline{(f \triangle g)^{(-1)}}\Pi_0 \\ &= \Pi_1(f \triangle g)\overline{(f \triangle g)^{(-1)}}\overline{\Pi_0} = \overline{\Pi_1(f \triangle g)\Pi_0} \Pi_1(f \triangle g) = \\ &\quad \overline{\Pi_1\Pi_0(f \triangle g)} \Pi_1(f \triangle g) = \overline{0} \Pi_1(f \triangle g) = 0 \end{aligned}$$

Similarly, $f\hat{g} = 0$.

Recall that Π_0 and Π_1 are jointly epic. We have $\Pi_0(f \triangle g)\hat{f} = f\hat{f} = f = \Pi_0(f \triangle 0)$ and $\Pi_1(f \triangle g)\hat{f} = g\hat{f} = 0 = \Pi_1(f \triangle 0)$. Therefore, $(f \triangle g)\hat{f} = f \triangle 0$.
 Similarly, $(f \triangle g)\hat{g} = 0 \triangle g$.

(v) Calculating, we have

$$f \nabla g(h \oplus k)\Pi_0 = f \nabla g\Pi_0 h = fh \text{ and } f \nabla g(h \oplus k)\Pi_1 = f \nabla g\Pi_1 k = gk,$$

but this means that $f \nabla g(h \oplus k) = fh \nabla gk$ by the joint monic property of Π_0 , Π_1 .

(vi) The proof for this is dual to the previous item, and depends on the joint epic property of Π_0 and Π_1 .

(vii) We are given $f \nabla g$ exists, therefore $f = (f \nabla g)\Pi_0$ and $g = (f \nabla g)\Pi_1$. But this means $hf = h(f \nabla g)\Pi_0$ and $hg = h(f \nabla g)\Pi_1$, from which we may conclude $hf \nabla hg = h(f \nabla g)$ by the fact that Π_0 and Π_1 are jointly monic. The proof of $(f \triangle g)h = fh \triangle gh$ is similar.

(viii) We are given $\bar{f} = \bar{f} \nabla \bar{g}\Pi_0$. Therefore,

$$\bar{f} = \bar{f}^{(-1)} = \Pi_0^{(-1)}(\bar{f} \nabla \bar{g})^{(-1)} = \Pi_0(\bar{f} \nabla \bar{g})^{(-1)}.$$

Similarly, $\bar{g} = \Pi_1(\bar{f} \nabla \bar{g})^{(-1)}$. But this means $(\bar{f} \nabla \bar{g})^{(-1)} = \bar{f} \triangle \bar{g}$.

(ix) Note that from item [\(v\)](#), we know that $f \nabla g = \bar{f} \nabla \bar{g}(f \oplus g)$. We are given $f' \leq f$ and $g' \leq g$. This gives us $\bar{f}'f = f'$, $\bar{g}'g = g'$, $\bar{f}'\bar{f} = \bar{f}'$ and $\bar{g}'\bar{g} = \bar{g}'$. Consider the map $\bar{f} \nabla \bar{g}(\bar{f}' \oplus \bar{g}')(f \oplus g)$. Calculating, we see

$$\begin{aligned} \bar{f} \nabla \bar{g}(\bar{f}' \oplus \bar{g}')(f \oplus g) &= \bar{f} \nabla \bar{g}(\bar{f}' \oplus \bar{g}')(\bar{f}' \oplus \bar{g}')(f \oplus g) \\ &= \bar{f} \nabla \bar{g}(\bar{f}' \oplus \bar{g}')(f' \oplus g') \\ &= \bar{f}\bar{f}' \nabla \bar{g}\bar{g}'(f' \oplus g') \\ &= \bar{f}'\bar{f} \nabla \bar{g}'\bar{g}(f' \oplus g') \\ &= \bar{f}' \nabla \bar{g}'(f' \oplus g') \\ &= f' \nabla g'. \end{aligned}$$

(x) From our diagram for \triangle , we know that $f^{(-1)} = (f \triangle g)^{(-1)}\Pi_0$ and $g^{(-1)} = (f \triangle g)^{(-1)}\Pi_1$. As well, we know that $\Pi_0(f \triangle g) = f$ and $\Pi_1(f \triangle g) = g$. Therefore, we have:

$$\Pi_0(f \triangle g)(f \triangle g)^{(-1)}\Pi_0 = \bar{f} \text{ and } \Pi_1(f \triangle g)(f \triangle g)^{(-1)}\Pi_1 = \bar{g}.$$

As $f \perp_{\oplus} g$, we know that $fg^{(-1)} = f\hat{g}g^{(-1)} = 0g^{(-1)} = 0$ and therefore,

$$\Pi_0(f \triangle g)(f \triangle g)^{(-1)}\Pi_1 = 0 \text{ and } \Pi_1(f \triangle g)(f \triangle g)^{(-1)}\Pi_0 = 0.$$

By Corollary [3.5.21](#) this means $(f \triangle g)(f \triangle g)^{(-1)} = \bar{f} \oplus \bar{g}$.

(xi) As $(f \nabla g) \oplus (h \nabla k) \Pi_0 = (f \nabla g)$ and $(f \nabla g) \oplus (h \nabla k) \Pi_1 = (h \nabla k)$, we see that $(f \nabla g) \oplus (h \nabla k)$ satisfies the diagram for $(f \oplus h) \nabla (g \oplus k)$. Dually, as $\Pi_0(f \triangle g) \oplus (h \triangle k) = (f \triangle g)$ and $\Pi_1(f \triangle g) \oplus (h \triangle k) = (h \triangle k)$, $(f \triangle g) \oplus (h \triangle k)$ satisfies the diagram for $(f \oplus h) \triangle (g \oplus k)$.

□

Definition 3.5.24. Define $f \perp_{\oplus} g$ when $f, g : A \rightarrow B$ and both $f \nabla g$ and $f \triangle g$.

Lemma 3.5.25. *If \mathbb{X} is an inverse category with a restriction zero and a disjointness tensor \oplus then the relation \perp_{\oplus} is a disjointness relation.*

Proof. We need to show that \perp_{\oplus} satisfies the disjointness axioms. We will use [Dis.6'] in place of [Dis.6] and [Dis.7] as discussed in Lemma 3.5.2.

[Dis.1] We must show $f \perp_{\oplus} 0$. This follows immediately from Lemma 3.5.23, item (ii).

[Dis.2] Show $f \perp_{\oplus} g$ implies $\bar{f}g = 0$. This is a direct consequence of Lemma 3.5.23, item (iii).

[Dis.3] We require $f \perp_{\oplus} g, f' \leq f, g' \leq g$ implies $f' \perp_{\oplus} g'$. From Lemma 3.5.23, item (ix), we immediately have $f' \nabla g'$ exists. Using a similar argument to the proof of this item, we also have $f' \triangle g'$ exists and hence $f' \perp_{\oplus} g'$.

[Dis.4] Commutativity of \perp_{\oplus} follows from the symmetry of the two required diagrams, see Lemma 3.5.23, item (i).

[Dis.5] Show that if $f \perp_{\oplus} g$ then $hf \perp_{\oplus} hg$ for any map h . By Lemma 3.5.23, item (vii), we know that $hf \nabla hg$ exists. By item (vi), $(hf) \triangle (hg) = (h \oplus h)(f \triangle g)$ and therefore $hf \perp_{\oplus} hg$.

[Dis.6'] We need to show $f \perp_{\oplus} g$ if and only if $\bar{f} \perp_{\oplus} \bar{g}$ and $\hat{f} \perp_{\oplus} \hat{g}$. This follows directly from Lemma 3.5.23, items (v) and (vi), which give us $f \nabla g = \bar{f} \nabla \bar{g}(f \oplus g)$ and $f \triangle g = (f \oplus g)\hat{f} \triangle \hat{g}$, where the equalities hold if either side of the equation exists.

□

The operations ∇ and Δ are sufficient to define a disjointness relation on an inverse category. However, when we wish to extend this to a disjoint join, we run into problems when trying to prove [DJ.4]. Specifically, there is not enough information to show that $\perp_{\oplus}[f, g, h]$ implies $f \perp_{\oplus} (g \sqcup_{\oplus} h)$.

Therefore, we add one more assumption regarding our tensor in order to define disjointness.

Definition 3.5.26. Let \mathbb{X} be an inverse category with a disjointness tensor \oplus and a restriction zero. Consider diagrams [equation \(3.15\)](#) and [equation \(3.16\)](#).

$$\begin{array}{ccc}
 A & \xrightarrow{g} & X \oplus Z \\
 \searrow \alpha & & \downarrow \Pi_0 \\
 X \oplus Y \oplus Z & \xrightarrow{\Pi_{0,2}} & X \oplus Z \\
 \downarrow \Pi_{0,1} & & \downarrow \Pi_0 \\
 X \oplus Y & \xrightarrow{\Pi_0} & X
 \end{array}
 \quad (3.15)$$

$$\begin{array}{ccc}
 X \oplus Y & \xrightarrow{\Pi_{0,1}} & X \oplus Y \oplus Z \\
 \uparrow \Pi_0 & & \uparrow \Pi_{0,2} \\
 X & \xrightarrow{\Pi_0} & X \oplus Y
 \end{array}
 \quad (3.16)$$

Then \oplus is a *disjoint sum tensor* when the following two conditions hold:

- α exists if and only if $f\Pi_1 \nabla g\Pi_1$ exists;
- β exists if and only if $\Pi_1 h \Delta \Pi_1 k$ exists.

Lemma 3.5.27. Let \mathbb{X} be an inverse category with a disjoint sum tensor as in Definition 3.5.26 and we are given $f, g, h : A \rightarrow B$ with $\perp_{\oplus}[f, g, h]$. Then both $f \nabla (g \nabla h)$ and $f \Delta (g \Delta h)$ exist.

Proof. As all the maps are disjoint, we know that each pair's ∇ map and Δ maps exist. Consider the diagram

$$\begin{array}{ccccc}
 A & \xrightarrow{f} & B & & \\
 & \searrow^{g \nabla f} & \uparrow \Pi_1 & & \\
 & \alpha \dashrightarrow & B \oplus B & \xrightarrow{\Pi_{0,2}} & B \oplus B \\
 & \searrow^{g \nabla h} & \downarrow \Pi_{0,1} & & \downarrow \Pi_0 \\
 & & B \oplus B & \xrightarrow{\Pi_0} & B
 \end{array}$$

where we claim $\alpha = (g \nabla h) \nabla f$.

The lower part of the diagram commutes as it fulfills the conditions of Definition 3.5.26. The upper triangle of the diagram commutes by the definition of $g \nabla f$. Noting that $\Pi_{0,1} : B \oplus B \oplus B \rightarrow B \oplus B$ is the same map as $\Pi_0 : (B \oplus B) \oplus B \rightarrow (B \oplus B)$ and $\Pi_{0,2}\Pi_1 : B \oplus B \oplus B \rightarrow B \oplus B \rightarrow B$ is the same map as $\Pi_1 : (B \oplus B) \oplus B \rightarrow B$, we see α does make the ∇ diagram for $g \nabla h$ and f commute. Therefore by Lemma 3.5.23, $f \nabla (g \nabla h)$ exists and is equal to $\alpha c_{\oplus\{01,2\}}$.

A dual diagram and reasoning shows $f \Delta (g \Delta h)$ exists. \square

Lemma 3.5.28. *In an inverse category \mathbb{X} with a disjoint sum tensor, when $\perp_{\oplus}[f, g, h]$, then:*

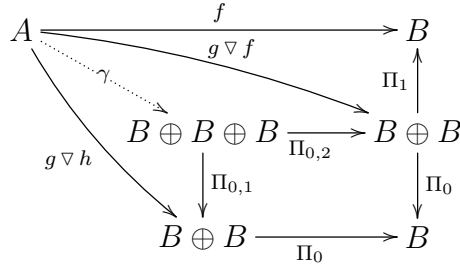
1. $f \nabla (g \nabla h) = ((f \nabla g) \nabla h) a_{\oplus}$ and both exist;
2. $f \Delta (g \Delta h) = ((f \Delta g) \Delta h) a_{\oplus}$ and both exist;

Proof. Consider the diagram

$$\begin{array}{ccccc}
 A & \xrightarrow{h} & B & & \\
 & \searrow^{f \nabla h} & \uparrow \Pi_1 & & \\
 & \alpha \dashrightarrow & B \oplus B \oplus B & \xrightarrow{\Pi_{0,2}} & B \oplus B \\
 & \searrow^{f \nabla g} & \downarrow \Pi_{0,1} & & \downarrow \Pi_0 \\
 & & B \oplus B & \xrightarrow{\Pi_0} & B
 \end{array}$$

which gives us $\alpha = (f \nabla g) \nabla h : A \rightarrow (B \oplus B) \oplus B$ and $\alpha a_{\oplus} : A \rightarrow B \oplus (B \oplus B)$. Next

consider the diagram



which gives us $\gamma c_{\oplus} = f \nabla (g \nabla h) : A \rightarrow B \oplus (B \oplus B)$.

Note from the diagrams we have:

$$\gamma c_{\oplus} \Pi_0 = f = \alpha a_{\oplus} \Pi_0$$

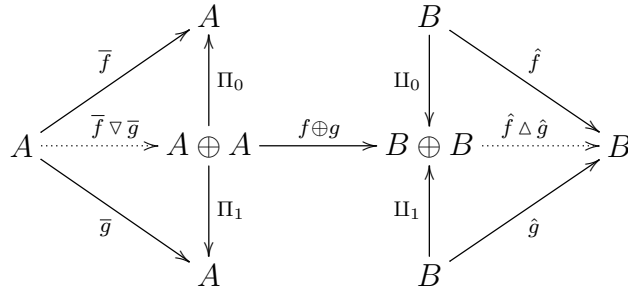
$$\gamma c_{\oplus} \Pi_1 \Pi_0 = g = \alpha a_{\oplus} \Pi_1 \Pi_0$$

$$\gamma c_{\oplus} \Pi_1 \Pi_1 = h = \alpha a_{\oplus} \Pi_1 \Pi_1.$$

Hence, by the assumption that Π_0, Π_1 are jointly monic, we have $\alpha = \gamma c_{\oplus} a_{\oplus}$ and hence $f \nabla (g \nabla h) = (f \nabla g) \nabla h$, up to the associativity isomorphism. \square

Definition 3.5.29. Let \mathbb{X} be an inverse category with a disjointness tensor and restriction zero. Assume we have two maps $f, g : A \rightarrow B$ with $f \perp_{\oplus} g$. Then define the map $f \sqcup_{\oplus} g = \bar{f} \nabla \bar{g} (f \oplus g) \hat{f} \triangle \hat{g}$.

For reference, the map $f \sqcup_{\oplus} g$ may be visualized as follows:



Using Lemma 3.5.23, we may rewrite this in a variety of equivalent ways:

$$\begin{aligned}
f \sqcup_{\oplus} g &= \bar{f} \nabla \bar{g} (f \oplus g) \hat{f} \Delta \hat{g} \\
&= f \nabla g \hat{f} \Delta \hat{g} \\
&= \bar{f} \nabla \bar{g} f \Delta g \\
&= f \nabla g (f^{(-1)} \oplus g^{(-1)}) f \Delta g
\end{aligned}$$

In particular, note that $\bar{f} \sqcup_{\oplus} \bar{g} = (\bar{f} \nabla \bar{g})(\bar{f} \Delta \bar{g})$ as $\hat{g} = \bar{g}$.

Lemma 3.5.30. *Let \mathbb{X} be an inverse category with a disjointness tensor and restriction zero.*

Let \mathbb{X} have the maps $f, g : A \rightarrow B$ with $f \perp_{\oplus} g$. Then \sqcup_{\oplus} has the following properties.

(i) *For all maps $h : A \rightarrow B$, $\bar{f}h \sqcup_{\oplus} \bar{g}h = (\bar{f} \sqcup_{\oplus} \bar{g})h$.*

(ii) *$\bar{f} \sqcup_{\oplus} \bar{g} = \overline{f \sqcup_{\oplus} g}$.*

Proof. (i) By Lemma 3.5.3, item (ii), we know that $\bar{f}h \perp_{\oplus} \bar{g}h$, hence we can form

$\bar{f}h \sqcup_{\oplus} \bar{g}h$. Also, noting that

$$\widehat{hfh} = \overline{hh^{(-1)}f} = \overline{hh^{(-1)}fh} = \overline{hfh} = \overline{fhh} = \overline{fh},$$

we may then calculate from the left hand side as follows:

$$\begin{aligned}
\bar{f}h \sqcup_{\oplus} \bar{g}h &= (\bar{f}h \nabla \bar{g}h)(\widehat{\bar{f}h} \Delta \widehat{\bar{g}h}) \\
&= (\bar{f} \nabla \bar{g})(\widehat{hfh} \Delta \widehat{hgh}) \\
&= (\bar{f} \nabla \bar{g})(\bar{f}h \Delta \bar{g}h) \\
&= (\bar{f} \nabla \bar{g})(\bar{f} \Delta \bar{g})h \\
&= (\bar{f} \sqcup_{\oplus} \bar{g})h.
\end{aligned}$$

(ii) Using Lemma 3.5.23, item (x), we can compute:

$$\begin{aligned}
\overline{f \sqcup_{\oplus} g} &= f \sqcup_{\oplus} g (f \sqcup_{\oplus} g)^{(-1)} \\
&= ((\bar{f} \nabla \bar{g})(f \triangle g)) \left((f \nabla g)^{(-1)} (\bar{f} \nabla \bar{g})^{(-1)} \right) \\
&= \bar{f} \nabla \bar{g} (f \nabla g) (f \nabla g)^{(-1)} \bar{f} \triangle \bar{g} \\
&= \bar{f} \nabla \bar{g} (\bar{f} \oplus \bar{g}) \bar{f} \triangle \bar{g} \\
&= \bar{f} \nabla \bar{g} \bar{f} \triangle \bar{g} \\
&= \bar{f} \sqcup_{\oplus} \bar{g}
\end{aligned}$$

□

Proposition 3.5.31. *Let \mathbb{X} be an inverse category with a disjoint sum tensor and restriction zero. Assume we have two maps f, g with $f \perp_{\oplus} g$. Then the map $f \sqcup_{\oplus} g$ from Definition 3.5.29 is a disjoint join.*

Proof. [DJ.1] We must show $f, g \leq f \sqcup_{\oplus} g$. Computing,

$$\begin{aligned}
\bar{f} (\bar{f} \nabla \bar{g}) f \triangle g &= (\bar{f} \nabla \bar{g}) \Pi_0 (\bar{f} \nabla \bar{g}) f \triangle g \\
&= \overline{(\bar{f} \nabla \bar{g}) \Pi_0 (\bar{f} \nabla \bar{g})} f \triangle g \\
&= (\bar{f} \nabla \bar{g}) \overline{\Pi_0} f \triangle g \\
&= (\bar{f} \nabla \bar{g}) \Pi_0 \Pi_0 f \triangle g \\
&= ((\bar{f} \nabla \bar{g}) \Pi_0) (\Pi_0 (f \triangle g)) \\
&= \bar{f} f \\
&= f
\end{aligned}$$

we see $f \leq f \sqcup_{\oplus} g$. Showing $g \leq f \sqcup_{\oplus} g$ proceeds in the same manner.

[DJ.2] We must show that $f \leq h$, $g \leq h$ and $f \perp_{\oplus} g$ implies $f \sqcup_{\oplus} g \leq h$. First, note that

$$\begin{aligned}
\overline{f \sqcup_{\oplus} g} h &= \overline{\overline{f} h \sqcup_{\oplus} \overline{g} h} h \\
&= \overline{(\overline{f} \sqcup_{\oplus} \overline{g}) h} h \\
&= \overline{(\overline{f} \sqcup_{\oplus} \overline{g}) h} h \\
&= \overline{(\overline{f} \sqcup_{\oplus} \overline{g}) h} (\overline{f} \sqcup_{\oplus} \overline{g}) h \\
&= \overline{(\overline{f} \sqcup_{\oplus} \overline{g}) h} (\overline{f} \sqcup_{\oplus} \overline{g}) h \\
&= (\overline{f} \sqcup_{\oplus} \overline{g}) h \\
&= (\overline{f} h \sqcup_{\oplus} \overline{g} h) \\
&= (f \sqcup_{\oplus} g)
\end{aligned}$$

[DJ.3] We must show stability of \sqcup_{\oplus} , i.e., that $h(f \sqcup_{\oplus} g) = hf \sqcup_{\oplus} hg$.

$$\begin{aligned}
h(f \sqcup_{\oplus} g) &= h((\overline{f} \nabla \overline{g})(f \triangle g)) \\
&= (h\overline{f} \nabla h\overline{g})(f \triangle g) \\
&= (\overline{h} f h \nabla \overline{h} g h)(f \triangle g) \\
&= (\overline{h} f \nabla \overline{h} g)(h \oplus h)(f \triangle g) \\
&= (\overline{h} f \nabla \overline{h} g)(hf \triangle hg) \\
&= hf \sqcup_{\oplus} hg
\end{aligned}$$

[DJ.4] We need to show $\perp_{\oplus}[f, g, h]$ if and only if $f \perp_{\oplus} (g \sqcup_{\oplus} h)$. For the right to left implication, note that the existence of $g \sqcup_{\oplus} h$ implies $g \perp_{\oplus} h$. We also know $g, h \leq g \sqcup_{\oplus} h$ by item 1 of this lemma. This gives us that $f \perp_{\oplus} g$ and $f \perp_{\oplus} h$, hence $\perp_{\oplus}[f, g, h]$. For the left to right implication, we use Lemma 3.5.27. As we have $\perp_{\oplus}[f, g, h]$, we know $f \nabla (g \nabla h)$ and $f \triangle (g \triangle h)$.

Recall that $g \sqcup_{\oplus} h = (g \nabla h)(\hat{g} \triangle \hat{h})$. Then the map

$$A \xrightarrow{f \nabla (g \nabla h)} B \oplus B \oplus B \xrightarrow{1 \oplus (\hat{g} \triangle \hat{h})} B \oplus B$$

makes the diagram for $f \nabla (g \sqcup_{\oplus} h)$ commute.

Recalling that $g \sqcup_{\oplus} h = (\bar{g} \nabla \bar{h})(g \triangle h)$, we also see that

$$A \oplus A \xrightarrow{1 \oplus (\bar{g} \nabla \bar{h})} A \oplus A \oplus A \xrightarrow{f \triangle (g \triangle h)} B$$

provides the witness map for $f \triangle (g \sqcup_{\oplus} h)$ and hence $f \perp_{\oplus} (g \sqcup_{\oplus} h)$.

□

3.6 Inverse sum categories

3.6.1 Inverse sums

Definition 3.6.1. In an inverse category with disjoint joins, an object X is the *inverse sum* of A and B when there exist maps i_0, i_1, x_0, x_1 such that:

- (i) i_0 and i_1 are monic;
- (ii) $i_0 : A \rightarrow X, i_1 : B \rightarrow X, x_0 : X \rightarrow A$ and $x_1 : X \rightarrow B$.
- (iii) $i_0^{(-1)} = x_0$ and $i_1^{(-1)} = x_1$.
- (iv) $i_0^{(-1)} i_0 \perp i_1^{(-1)} i_1$ and $i_0^{(-1)} i_0 \sqcup i_1^{(-1)} i_1 = 1_X$.

i_0 and i_1 will be referred to as the *injection* maps of the inverse sum.

Lemma 3.6.2. *The inverse sum X of A and B is unique up to isomorphism.*

Proof. Assume we have two inverse sums over A and B :

$$A \begin{array}{c} \xrightarrow{i_0} \\ \xleftarrow{x_0} \end{array} X \begin{array}{c} \xleftarrow{i_1} \\ \xrightarrow{x_1} \end{array} B \quad \text{and} \quad A \begin{array}{c} \xrightarrow{j_0} \\ \xleftarrow{y_0} \end{array} Y \begin{array}{c} \xleftarrow{j_1} \\ \xrightarrow{y_1} \end{array} B.$$

We will show that the map $x_0 j_0 \sqcup x_1 j_1 : X \rightarrow Y$ is an isomorphism.

Note by the fact that i_1 is monic, we may conclude from the definition that $0 = \overline{x_0 i_0 x_1}$ and therefore $0 = x_0 i_0 x_1$. Then, given that x_0 is the inverse of the monic i_0 , we may calculate $0 = \hat{0} = \widehat{x_0 i_0 x_1} = \overline{x_1^{(-1)} i_0^{(-1)} i_0} = \overline{x_1^{(-1)} i_0^{(-1)}} = \widehat{i_0 x_1}$. From this we see $i_0 x_1 = 0$. Similarly, we have $i_1 x_0 = 0, j_0 y_1 = 0$ and $j_1 y_0 = 0$.

Next, by Lemma 3.5.3, we know that $\overline{x_0} \perp \overline{x_1}$ as both i_0 and i_1 are monic. By the same lemma, $\hat{j}_0 \perp \hat{j}_1$ as y_0, y_1 are the inverses of monic maps. Then, from [Dis.7], we have $x_0 j_0 \perp x_1 j_1$, hence we may form $x_0 j_0 \sqcup x_1 j_1 : X \rightarrow Y$.

Similarly, we may form the map $y_0 i_0 \sqcup y_1 i_1 : Y \rightarrow X$. Computing their composition:

$$\begin{aligned}
(x_0 j_0 \sqcup x_1 j_1)(y_0 i_0 \sqcup y_1 i_1) &= (x_0 j_0(y_0 i_0 \sqcup y_1 i_1)) \sqcup (x_1 j_1(y_0 i_0 \sqcup y_1 i_1)) \\
&= x_0 j_0 y_0 i_0 \sqcup x_0 j_0 y_1 i_1 \sqcup x_1 j_1 y_0 i_0 \sqcup x_1 j_1 y_1 i_1 \\
&= x_0 1 i_0 \sqcup x_0 0 i_1 \sqcup x_1 0 i_0 \sqcup x_1 1 i_1 \\
&= x_0 i_0 \sqcup x_1 i_1 = 1.
\end{aligned}$$

Computing the other direction,

$$\begin{aligned}
(y_0 i_0 \sqcup y_1 i_1)(x_0 j_0 \sqcup x_1 j_1) &= (y_0 i_0(x_0 j_0 \sqcup x_1 j_1)) \sqcup (y_1 i_1(x_0 j_0 \sqcup x_1 j_1)) \\
&= y_0 i_0 x_0 j_0 \sqcup y_0 i_0 x_1 j_1 \sqcup y_1 i_1 x_0 j_0 \sqcup y_1 i_1 x_1 j_1 \\
&= y_0 1 j_0 \sqcup y_0 0 j_1 \sqcup y_1 0 j_0 \sqcup y_1 1 j_1 \\
&= y_0 j_0 \sqcup y_1 j_1 = 1.
\end{aligned}$$

This shows that the map between any two inverse sums is an isomorphism.

□

Definition 3.6.3. Suppose \mathbb{X} is an inverse category with with disjoint joins \sqcup based on a disjointness relation \perp and a restriction zero. If every pair of objects has an inverse sum as in Definition 3.6.1, we call the category an *inverse sum* category. For any two objects A, B in \mathbb{X} , we write their inverse sum as $A + B$.

Lemma 3.6.4. *Let \mathbb{X} be an inverse category with a restriction 0 and a disjoint sum tensor \oplus . Then \mathbb{X} is an inverse sum category.*

Proof. We claim that setting $i_i = \Pi_i$ and $x_i = \Pi_i$ and setting $X = A \oplus B$ produces inverse sums in \mathbb{X} and show this satisfies the four conditions of Definition 3.6.1.

- (i) From Lemma 3.5.20, we know that Π_0 and Π_1 are monic maps.
- (ii) $\Pi_0 : A \rightarrow A \oplus B$, $\Pi_1 : B \rightarrow A \oplus B$, $\Pi_0 : A \oplus B \rightarrow A$ and $\Pi_1 : A \oplus B \rightarrow B$.
- (iii) $\Pi_0^{(-1)} = \Pi_0$ and $\Pi_1^{(-1)} = \Pi_1$.
- (iv) $i_0^{(-1)}i_0 = 1 \oplus 0 \perp_{\oplus} 0 \oplus 1 = i_1^{(-1)}i_1$ as $1 \oplus 0 \nabla 0 \oplus 1 = (u_{\oplus}^{r(-1)} \oplus u_{\oplus}^{l(-1)})$ and $1 \oplus 0 \triangle 0 \oplus 1 = (\Pi_0 \oplus \Pi_1)$. For their join, $(1 \oplus 0) \sqcup_{\oplus} (0 \oplus 1) = (u_{\oplus}^{r(-1)} \oplus u_{\oplus}^{l(-1)})(\Pi_0 \oplus \Pi_1) = u_{\oplus}^{r(-1)}\Pi_0 \oplus u_{\oplus}^{l(-1)}\Pi_1 = 1 \oplus 1 = 1$.

□

Lemma 3.6.5. *If A is an object in \mathbb{X} , an inverse sum category, then $A + 0$ is isomorphic to A .*

Proof. We write the inverse sum diagram:

$$A \begin{array}{c} \xrightarrow{1} \\ \xleftarrow{1} \end{array} A \begin{array}{c} \xleftarrow{0} \\ \xrightarrow{0} \end{array} 0.$$

□

Lemma 3.6.6. *Suppose \mathbb{X} is an inverse sum category and \mathbb{Y} is an inverse category with a restriction zero. Further, suppose $F : \mathbb{X} \rightarrow \mathbb{Y}$ is a restriction functor which preserves disjoint joins. Then, F preserves inverse sums.*

Proof. In \mathbb{X} , consider the inverse sum over A and B ,

$$A \begin{array}{c} \xrightarrow{i_0} \\ \xleftarrow{x_0} \end{array} X \begin{array}{c} \xleftarrow{i_1} \\ \xrightarrow{x_1} \end{array} B.$$

The functor F maps this as follows:

$$F(A) \begin{array}{c} \xrightarrow{F(i_0)} \\ \xleftarrow{F(x_0)} \end{array} F(X) \begin{array}{c} \xleftarrow{F(i_1)} \\ \xrightarrow{F(x_1)} \end{array} F(B).$$

As F is a restriction functor, we immediately have $F(x_0) = F(i_0^{(-1)}) = F(i_0)^{(-1)}$ and $F(x_1) = F(i_1)^{(-1)}$. Since F preserves the disjoint join, we also have $F(i_0)^{(-1)}F(i_0) \perp F(i_1)^{(-1)}F(i_1)$ and $F(i_0)^{(-1)}F(i_0) \sqcup F(i_1)^{(-1)}F(i_1) = 1$.

Finally, as F is a restriction functor, it preserves monics, hence $F(i_0)$ and $F(i_1)$ are both monic and therefore $F(X)$ is the inverse sum of $F(A)$ and $F(B)$.

□

Lemma 3.6.7. *Given \mathbb{X} an inverse sum category and maps $f : A \rightarrow C$ and $g : B \rightarrow D$ in \mathbb{X} . Then $i_0^{(-1)}f i_0 \perp i_1^{(-1)}g i_1 : A + B \rightarrow A + B$.*

Proof. Note that $\overline{i_0^{(-1)}f i_0} = \overline{i_0^{(-1)}f} \leq \overline{i_0^{(-1)}}$ and similarly $\overline{i_1^{(-1)}g i_1} \leq \overline{i_1^{(-1)}}$. Then, by [Dis.3], we have $\overline{i_0^{(-1)}f i_0} \perp \overline{i_1^{(-1)}g i_1}$.

Then, as $\widehat{i_0^{(-1)}f i_0} \leq \widehat{i_0}$ and $\widehat{i_1^{(-1)}g i_1} \leq \widehat{i_1}$, this means we have $\widehat{i_0^{(-1)}f i_0} \perp \widehat{i_1^{(-1)}g i_1}$ and by Lemma 3.5.3, this means $i_0^{(-1)}f i_0 \perp i_1^{(-1)}g i_1$. □

Lemma 3.6.8. *Given \mathbb{X} is an inverse sum category. Denote the inverse sum of objects A, B of \mathbb{X} by $A + B$. Then for objects A, B and X with maps $f : A \rightarrow X$ and $g : B \rightarrow X$ such that $\hat{f} \perp \hat{g}$, there exists a unique map h making the following diagram commute.*

$$\begin{array}{ccc}
 A & & \\
 \downarrow i_0 & \searrow f & \\
 A + B & \xrightarrow{\quad h \quad} & X \\
 \uparrow i_1 & \nearrow g & \\
 B & &
 \end{array}$$

We use the notation $f \hat{+} g$ for the unique map h .

Proof. As $\hat{f} \perp \hat{g}$ and $\overline{i_0^{(-1)}} \perp \overline{i_1^{(-1)}}$ we may form the map $h' = i_0^{(-1)}f \sqcup i_1^{(-1)}g$. By its construction, h' is a map from $A + B$ to X which makes the diagram commute. Suppose now that both maps v and w are such maps. Then we have

$$(i_0^{(-1)}i_0)v = (i_0^{(-1)}i_0)w \quad \text{and} \quad (i_1^{(-1)}i_1)v = (i_1^{(-1)}i_1)w.$$

As $i_0^{(-1)}i_0 \perp i_1^{(-1)}i_1$, by Lemmas 3.5.3 and 3.5.13, we know that $(i_0^{(-1)}i_0)v \perp (i_1^{(-1)}i_1)v$ and $(i_0^{(-1)}i_0)w \perp (i_1^{(-1)}i_1)w$ allowing us to form their respective disjoint joins. As the disjoint

joins of equal maps remains equal, we have

$$(i_0^{(-1)}i_0)v \sqcup (i_1^{(-1)}i_1)v = (i_0^{(-1)}i_0)w \sqcup (i_1^{(-1)}i_1)w$$

$$(i_0^{(-1)}i_0 \sqcup i_1^{(-1)}i_1)v = (i_0^{(-1)}i_0 \sqcup i_1^{(-1)}i_1)w$$

$$(1)v = (1)w$$

$$v = w.$$

□

Corollary 3.6.9. *Given \mathbb{X} is an inverse sum category. Then for objects A, B and X with maps $f : X \rightarrow A$ and $g : X \rightarrow B$ such that $\bar{f} \perp \bar{g}$, there exists a unique map h making the following diagram commute.*

$$\begin{array}{ccc} & & A \\ & \nearrow f & \downarrow i_0 \\ X & \xrightarrow{\quad h \quad} & A + B \\ & \searrow g & \uparrow i_1 \\ & & B. \end{array}$$

We use the notation $f \bar{+} g$ for the unique map h .

Proof. This is simply the dual of Lemma 3.6.8. The unique map h in this case is $fi_0 \sqcup gi_1$. □

Corollary 3.6.10. *Suppose \mathbb{X} is an inverse sum category. Then for objects A, B, C and D with maps $f : A \rightarrow C$ and $g : B \rightarrow D$, there exists a unique map h making the following diagram commute.*

$$\begin{array}{ccc} A & \xrightarrow{\quad f \quad} & C \\ i_0 \downarrow & & \downarrow i_0 \\ A + B & \xrightarrow{\quad h \quad} & C + D \\ i_1 \uparrow & & \uparrow i_1 \\ B & \xrightarrow{\quad g \quad} & D. \end{array}$$

We use the notation $f + g$ for the map h .

Proof. This follows directly from Lemma 3.6.8 by setting $X = C + D$. The unique map in this case is $i_0^{(-1)} f i_0 \sqcup i_1^{(-1)} g i_1$. \square

Lemma 3.6.11. *Suppose \mathbb{X} and \mathbb{Y} are inverse sum categories and $F : \mathbb{X} \rightarrow \mathbb{Y}$ is a restriction functor which preserves inverse sums. Then, F preserves disjoint joins.*

Proof. By stating that F preserves the inverse sum, we mean it preserves diagrams derived via the properties of the inverse sum, and specifically, it will preserve the diagrams of Lemma 3.6.8 and Corollaries 3.6.9 and 3.6.10.

Suppose we are given $f, g : A \rightarrow B$ with $f \perp g$. In the inverse sum category, we know that $f \sqcup g = (\bar{f} i_0 \sqcup \bar{g} i_1)(i_0^{(-1)} f i_0 \sqcup i_1^{(-1)} g i_1)(i_0^{(-1)} \hat{f} \sqcup i_1^{(-1)} \hat{g})$, as this follows by:

1. Apply Corollary 3.6.9 to \bar{f} and \bar{g} ;
2. then apply Corollary 3.6.10 to f, g ;
3. finally apply Lemma 3.6.8 to \hat{f}, \hat{g} .

In the notation above, we have that $f \sqcup g = (\bar{f} \bar{+} \bar{g})(f + g)(\hat{f} \hat{+} \hat{g})$. This gives us, as F preserves the inverse sum:

$$\begin{aligned}
F(f \sqcup g) &= F(\bar{f} \bar{+} \bar{g}) F(f + g) F(\hat{f} \hat{+} \hat{g}) \\
&= (F(\bar{f}) \bar{+} F(\bar{g}))(F(f) + F(g))(F(\hat{f}) \hat{+} F(\hat{g})) \\
&= (\overline{F(f)} \bar{+} \overline{F(g)})(F(f) + F(g))(\widehat{F(f)} \hat{+} \widehat{F(g)}) \\
&= F(f) \sqcup F(g).
\end{aligned}$$

The last line in the above is due to \mathbb{Y} being an inverse sum category as well. \square

3.6.2 Inverse sum tensor

Inverse sum tensor definitions

Definition 3.6.12. An *inverse sum tensor* in an inverse category \mathbb{X} with disjoint joins \sqcup based on a disjointness relation \perp and a restriction zero is given by a tensor combined with

two restriction monics, Π_0 and Π_1 . The data for the tensor is:

$$_-\oplus_- : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{X} \quad (\text{a restriction functor preserving disjoint joins})$$

$$0 : \mathbf{1} \rightarrow \mathbb{X}$$

$$u_\oplus^l : 0 \oplus A \rightarrow A$$

$$u_\oplus^r : A \oplus 0 \rightarrow A$$

$$a_\oplus : (A \oplus B) \oplus C \rightarrow A \oplus (B \oplus C)$$

$$c_\oplus : A \oplus B \rightarrow B \oplus A$$

$$\Pi_0 : A \rightarrow A \oplus B$$

$$\Pi_1 : B \rightarrow A \oplus B$$

where $u_\oplus^l, u_\oplus^r, a_\oplus, c_\oplus$ are all isomorphisms and the standard symmetric monoidal equations and coherence diagrams hold. The unit of the tensor, $0 : \mathbf{1} \rightarrow \mathbb{X}$, is the restriction zero of the category. We specifically note that preserving disjoint joins means the tensor obeys the following two equations:

$$f \perp g, h \perp k \text{ implies } f \oplus h \perp g \oplus k \tag{3.17}$$

$$f \perp g, h \perp k \text{ implies } (f \sqcup g) \oplus (h \sqcup k) = (f \oplus h) \sqcup (g \oplus k). \tag{3.18}$$

Disjointness and the inverse sum tensor

Lemma 3.6.13. *Given an inverse category \mathbb{X} with a disjoint sum tensor \oplus as in Definition 3.5.26, then \oplus is an inverse sum tensor.*

Proof. From the data of the disjoint sum tensor, the only thing remaining to show is that the tensor preserves the disjoint join.

Suppose we have $f \perp_\oplus g$ and $h \perp_\oplus k$. From Lemma 3.5.23, item (xi), we know both $(f \oplus h) \nabla (g \oplus k)$ and $(f \oplus h) \Delta (g \oplus k)$ exist, hence $(f \oplus h) \perp_\oplus (g \oplus k)$. This shows condition (equation (3.17)).

For condition (equation (3.18)), we compute from the right hand side:

$$\begin{aligned}
(f \oplus h) \sqcup_{\oplus} (g \oplus k) &= (f \oplus h) \nabla (g \oplus k) (\widehat{f \oplus h}) \Delta (\widehat{g \oplus k}) \\
&= ((f \nabla g) \oplus (h \nabla k)) \left((\hat{f} \oplus \hat{h}) \Delta (\hat{g} \oplus \hat{k}) \right) \\
&= ((f \nabla g) \oplus (h \nabla k)) \left((\hat{f} \Delta \hat{g}) \oplus (\hat{h} \Delta \hat{k}) \right) \\
&= \left((f \nabla g) (\hat{f} \Delta \hat{g}) \right) \oplus \left((h \nabla k) (\hat{h} \Delta \hat{k}) \right) \\
&= (f \sqcup_{\oplus} g) \oplus (h \sqcup_{\oplus} k).
\end{aligned}$$

The second and third lines above again use Lemma 3.5.23, item (xi). □

Inverse sums and the inverse sum tensor

Lemma 3.6.14. *If \oplus is an inverse sum tensor in the inverse category \mathbb{X} , then $A \oplus B \cong A + B$, an inverse sum of A and B .*

Proof. As \oplus is a restriction functor from $\mathbb{X} \times \mathbb{X}$ to \mathbb{X} , this actually follows immediately from Lemma 3.6.6. It may also be proven directly:

Draw the inverse sum diagram:

$$\begin{array}{ccccc}
A & \xrightarrow{i_0 = u_{\oplus}^r (-1) (1 \oplus 0)} & A \oplus B & \xleftarrow{i_1 = u_{\oplus}^l (-1) (0 \oplus 1)} & B \\
& \xleftarrow{x_0 = (1 \oplus 0) u_{\oplus}^r} & & \xrightarrow{x_1 = (0 \oplus 1) u_{\oplus}^l} &
\end{array}$$

Therefore, we have $i_0^{(-1)} i_0 = (1 \oplus 0) u_{\oplus}^r u_{\oplus}^r (-1) (1 \oplus 0) = (1 \oplus 0) (1 \oplus 0) = (1 \oplus 0)$. Similarly, $i_1^{(-1)} i_1 = (0 \oplus 1)$. Since $0 \perp 1$, we have $i_0^{(-1)} i_0 \perp i_1^{(-1)} i_1$.

By the functoriality of \oplus and that it preserves disjoint joins, we have $(1 \oplus 0) \sqcup (0 \oplus 1) = (1 \sqcup 0) \oplus (0 \sqcup 1) = 1 \oplus 1 = 1_{A \oplus B}$. Hence $A \oplus B$ is an inverse sum of A and B and by Lemma 3.6.2 it is isomorphic to $A + B$. □

Conversely, we can show that given a tensor which produces inverse sums, that tensor will be an inverse sum tensor.

Lemma 3.6.15. *Given an inverse category \mathbb{X} with restriction zero, a disjointness relation \perp , a disjoint join \sqcup and a symmetric monoidal tensor \oplus , with natural restriction monics $\Pi_0 : A \rightarrow A \oplus B$ and $\Pi_1 : B \rightarrow A \oplus B$ such that $A \oplus B$ is an inverse sum under Π_0 and Π_1 , then when $f, g : A \rightarrow B$ and $h, k : C \rightarrow D$ with $f \perp g$ and $h \perp k$, then $f \oplus h \perp g \oplus k$ and $(f \oplus h) \sqcup (g \oplus k) = (f \sqcup g) \oplus (h \sqcup k)$.*

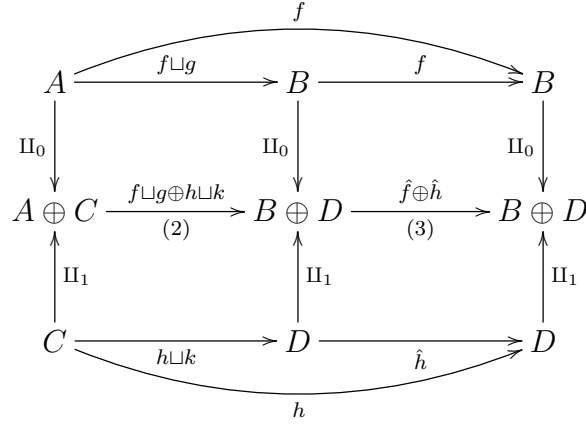
Proof. Similarly, this follows immediately from Lemma 3.6.11. We show it directly below:

$$\begin{array}{ccccc}
 & & f & & \\
 & \xrightarrow{\quad \bar{f} \quad} & A & \xrightarrow{\quad f \sqcup g \quad} & B \\
 \Pi_0 \downarrow & & \downarrow \Pi_0 & & \downarrow \Pi_0 \\
 A \oplus C & \xrightarrow[\quad (1) \quad]{\quad \bar{f} \oplus \bar{h} \quad} & A \oplus C & \xrightarrow[\quad (2) \quad]{\quad f \sqcup g \oplus h \sqcup k \quad} & B \oplus D \\
 \uparrow \Pi_1 & & \uparrow \Pi_1 & & \uparrow \Pi_1 \\
 C & \xrightarrow[\quad \bar{h} \quad]{} & C & \xrightarrow[\quad h \sqcup k \quad]{} & D \\
 & & h & &
 \end{array}$$

Consider $\Pi_0^{(-1)} \bar{f} \Pi_0$. As this is idempotent and we are in an inverse category, we know that $\Pi_0^{(-1)} \bar{f} \Pi_0 = \overline{\Pi_0^{(-1)} \bar{f} \Pi_0} = \overline{\Pi_0^{(-1)} \bar{f}} = \widehat{\bar{f} \Pi_0}$. Similarly, $\Pi_1^{(-1)} \bar{h} \Pi_1 = \widehat{\bar{h} \Pi_1}$. By [Dis.5] and [Dis.6], we know that $\widehat{\bar{f} \Pi_0} \perp \widehat{\bar{g} \Pi_0}$ and $\widehat{\bar{h} \Pi_1} \perp \widehat{\bar{k} \Pi_1}$. Additionally, as shown in the proof of Lemma 3.6.2, we know $\widehat{\Pi_0} \perp \widehat{\Pi_1}$. Hence, by [Dis.3], we have $\widehat{\bar{f} \Pi_0} \perp \widehat{\bar{g} \Pi_0}$ for any maps x, y .

Hence, we can form the map $\widehat{\bar{f} \Pi_0} \sqcup \widehat{\bar{h} \Pi_1}$. Referring to the commutative diagram above, by Corollary 3.6.10 there is an unique map at location (1) which makes the diagram commute — currently given as $\bar{f} \oplus \bar{h}$. But, the map $\widehat{\bar{f} \Pi_0} \sqcup \widehat{\bar{h} \Pi_1}$ also satisfies this, hence we have $\widehat{\bar{f} \Pi_0} \sqcup \widehat{\bar{h} \Pi_1} = \bar{f} \oplus \bar{h}$. Similarly, $\widehat{\bar{g} \Pi_0} \sqcup \widehat{\bar{k} \Pi_1} = \bar{g} \oplus \bar{k}$. But, by Lemma 3.5.15, this means $\overline{\bar{f} \oplus \bar{h}} \perp \overline{\bar{g} \oplus \bar{k}}$.

Using a similar argument based on the diagram



we can show $\widehat{f \oplus h} \perp \widehat{g \oplus k}$ and therefore $f \oplus h \perp g \oplus k$.

This allows us to form the map $(f \oplus h) \sqcup (g \oplus k)$. Once again, as the objects are inverse sums, the map at (2) is unique. However, we see that both $f \sqcup g \oplus h \sqcup k$ and $(f \oplus h) \sqcup (g \oplus k)$ fulfill this requirement and hence they are equal. \square

Definition 3.6.16. An inverse category \mathbb{X} with restriction zero, a disjointness relation \perp , a disjoint join \sqcup and an inverse sum tensor \oplus is called an *inverse sum tensor category*.

Corollary 3.6.17. In an inverse sum tensor category, $f \oplus g$ is given by $i_0^{(-1)} f i_0 \sqcup i_1^{(-1)} g i_1$.

Proof. Recall that in the proof of Lemma 3.6.2 that we showed $\overline{i_0^{(-1)}} \perp \overline{i_1^{(-1)}}$ and $\hat{i}_0 \perp \hat{i}_1$. Hence, by [Dis.7], we know that $i_0^{(-1)} f i_0 \perp i_1^{(-1)} g i_1$ and we can therefore form the disjoint join. \square

Matrices

In this section, we will show that when given an inverse sum category \mathbb{X} , one can define a type of matrix category based on \mathbb{X} . We will call this category $iMat(\mathbb{X})$. Furthermore, we will show that $iMat(\mathbb{X})$ has is an inverse category and that \mathbb{X} embeds within this category.

Definition 3.6.18. A matrix of maps $[f_{ij}]$ in an inverse sum category \mathbb{X} which satisfy the condition:

$$\text{For each } i, \text{ whenever } j \neq k, \text{ then } f_{ij} \perp f_{ik} \quad (3.19)$$

is called an *inverse sum matrix*.

In the following we will use the notation Π_i for the i^{th} injection map of the inverse sum.

Further, we define a monoid \diamond of any two such matrices by $[h_{ik}] = [f_{ij}] \diamond [g_{jk}]$ where the element h_{ik} of the matrix is given by:

$$h_{ik} = \bigsqcup_j f_{ij} \Pi_i g_{jk}$$

noting again that composition of maps is diagrammatic order. Of course, this is only defined when the domain of g_{jk} is the inverse sum of some objects in \mathbb{X} and the range of f_{ij} .

We also define \diamond for an inverse sum matrix $[f_{ij}]$ and a row (list) of objects $[A_i]$ in \mathbb{X} by $[B_j] = [A_i] \diamond [f_{ij}]$ is given by

$$B_j = \Sigma_i f_{ij} A_i$$

where Σ stands for the inverse sum operation of the category \mathbb{X} . Note this requires choosing a specific inverse sum for each pair of objects from the set of isomorphic inverse sums.

Definition 3.6.19. Given an inverse sum category \mathbb{X} , we define $iMat(\mathbb{X})$, the *inverse matrix category* of \mathbb{X} as follows:

Objects: Lists of the objects of \mathbb{X} .

Maps: The inverse sum matrix $[f_{ij}] : [A_i] \rightarrow [B_j]$ where each individual map $f_{ij} : A_i \rightarrow B'_{ij}$ is a map in \mathbb{X} . The B_j are the chosen inverse sums of the B'_{ij} . The result of $[f_{ij}][A_i]$ is given by $[A_i] \diamond [f_{ij}]$.

Identity: The inverse sum matrix I .

Composition: Given $[f_{ij}] : [A_i] \rightarrow [B_j]$ and $[g_{jk}] : [B_j] \rightarrow [C_k]$, then $[f_{ij}][g_{jk}] : [A_i] \rightarrow [C_k]$ is defined as $[g_{jk}] \diamond [f_{ij}]$.

Restriction: We set $\overline{[f_{ij}]}$ to be $[f'_{ij}]$ where $f'_{ij} = 0$ when $i \neq j$ and $f'_{ii} = \sqcup_j \overline{f_{ij}}$.

We will use the notation $\delta[d_1, d_2, \dots, d_n]$ for a diagonal $n \times n$ matrix with entries along the diagonal of $[d_1, d_2, \dots, d_n]$ and $\delta_j[d_j]$ for diagonal matrices where the j, j entry is d_j .

Lemma 3.6.20. *When \mathbb{X} is an inverse sum category, $iMat(\mathbb{X})$ is a restriction category.*

Proof. We need to show the following:

- Composition is well defined and associative;
- The restriction is well defined.

Composition is well defined: Consider $[h_{ik}] = [f_{ij}][g_{jk}] (= [g_{jk}] \diamond [f_{ij}])$ where

$$[f_{ij}] : [A_1, \dots, A_n] \rightarrow [B_1, \dots, B_m] \text{ and } [g_{jk}] : [B_1, \dots, B_m] \rightarrow [C_1, \dots, C_\ell].$$

By Definition 3.6.18, we know $h_{ik} = \sqcup_j f_{ij} g_{jk}$. Individually, for each j we know the composition $f_{ij} \amalg_i g_{jk}$ is defined and is from A_i to C'_{jk} . By the stability and universality of \perp , we know h_{ik} exists and by the definition of \sqcup , we have each $h_{ik} : A_i \rightarrow C'_{jk}$ and hence composition is well-defined.

Associativity of composition. We have

$$\begin{aligned} ([f_{ij}][g_{jk}])[h_{k\ell}] &= \left[\left(\sqcup_j f_{ij} \amalg_i g_{jk} \right) \right] [h_{k\ell}] \\ &= \left[\sqcup_k \left(\sqcup_j f_{ij} \amalg_i g_{jk} \right) \amalg_j h_{k\ell} \right] \\ &= \left[\sqcup_j f_{ij} \amalg_i \left(\sqcup_k g_{jk} \amalg_j h_{k\ell} \right) \right] \\ &= [f_{ij}]([g_{jk}][h_{k\ell}]) \end{aligned}$$

For the restriction axioms:

$$[\mathbf{R.1}] \quad \overline{[f_{ij}]}[f_{ij}] = \begin{bmatrix} (\sqcup_j \overline{f_{1j}})f_{11} & \cdots & (\sqcup_j \overline{f_{1j}})f_{1n} \\ & \vdots & \\ (\sqcup_j \overline{f_{mj}})f_{m1} & \cdots & (\sqcup_j \overline{f_{mj}})f_{mn} \end{bmatrix} = [f_{ij}].$$

[R.2] $\overline{[f_{ij}]} \overline{g_{ij}} = \overline{g_{ij}} \overline{[f_{ij}]}$ as diagonal matrices commute and \sqcup is also commutative.

$$\begin{aligned}
[\mathbf{R.3}] \quad \overline{[f_{ij}][g_{jk}]} &= \overline{\delta[\sqcup_j \overline{f_{1j}}, \dots, \sqcup_j \overline{f_{nj}}][g_{jk}]} \\
&= \overline{\begin{bmatrix} \sqcup_j \overline{f_{1j}} g_{11} & \dots & \sqcup_j \overline{f_{1j}} g_{1k} \\ & \vdots & \\ \sqcup_j \overline{f_{nj}} g_{n1} & \dots & \sqcup_j \overline{f_{nj}} g_{nk} \end{bmatrix}} \\
&= \delta[\sqcup_k (\sqcup_j (\overline{f_{1j}} g_{1k})), \dots, \sqcup_k (\sqcup_j (\overline{f_{nj}} g_{nk}))] \\
&= \delta[\sqcup_k (\sqcup_j (\overline{f_{1j}}) \overline{g_{1k}}), \dots, \sqcup_k (\sqcup_j (\overline{f_{nj}}) \overline{g_{nk}})] \\
&= \delta[(\sqcup_j (\overline{f_{1j}}) \sqcup_k \overline{g_{1k}}), \dots, (\sqcup_j (\overline{f_{nj}}) \sqcup_k \overline{g_{nk}})] \\
&= \overline{[f_{ij}][g_{jk}]}
\end{aligned}$$

$$\begin{aligned}
[\mathbf{R.4}] \quad [f_{ij}][\overline{g_{jk}}] &= [f_{ij}]\delta_j[\sqcup_k \overline{g_{jk}}] \\
&= \begin{bmatrix} f_{11} \sqcup_k \overline{g_{1k}} & \cdots & f_{1n} \sqcup_k \overline{g_{nk}} \\ & \vdots & \\ f_{m1} \sqcup_k \overline{g_{1k}} & \cdots & f_{mn} \sqcup_k \overline{g_{nk}} \end{bmatrix} \\
&= \begin{bmatrix} \sqcup_k f_{11} \overline{g_{1k}} & \cdots & \sqcup_k f_{1n} \overline{g_{nk}} \\ & \vdots & \\ \sqcup_k f_{m1} \overline{g_{1k}} & \cdots & \sqcup_k f_{mn} \overline{g_{nk}} \end{bmatrix} \\
&= \begin{bmatrix} \sqcup_k \overline{f_{11} g_{1k}} f_{11} & \cdots & \sqcup_k \overline{f_{1n} g_{nk}} f_{1n} \\ & \vdots & \\ \sqcup_k \overline{f_{m1} g_{1k}} f_{m1} & \cdots & \sqcup_k \overline{f_{mn} g_{nk}} f_{mn} \end{bmatrix} \\
&= \begin{bmatrix} \sqcup_j \sqcup_k \overline{f_{1j} g_{jk}} f_{11} & \cdots & \sqcup_j \sqcup_k \overline{f_{1j} g_{jk}} f_{1n} \\ & \vdots & \\ \sqcup_j \sqcup_k \overline{f_{mj} g_{jk}} f_{m1} & \cdots & \sqcup_j \sqcup_k \overline{f_{mj} g_{jk}} f_{mn} \end{bmatrix} \\
&= \overline{[f_{ij}][g_{jk}]}[f_{ij}].
\end{aligned}$$

□

Lemma 3.6.21. *In an inverse sum tensor category, any map $f : A \oplus B \rightarrow C \oplus D$ may be represented in a matrix form. Composition of maps may be computed by multiplication of the matrices, with composition taking the place of base level multiplication and djoin the place of addition.*

Proof. Recall from Lemma 3.6.14 that $A \oplus B$ and $C \oplus D$ are inverse sums. Referencing Definition 3.6.1, define $e_0 = i_0^{(-1)}i_0$ and $e_1 = i_1^{(-1)}i_1$ and recall that $e_0 \perp e_1$, $e_0 \sqcup e_1 = 1$.

Then given a function $f : A \oplus B \rightarrow C \oplus D$ define

$$f_M = \begin{bmatrix} e_0 f e_0 & e_0 f e_1 \\ e_1 f e_0 & e_1 f e_1 \end{bmatrix}.$$

Note first that since $e_0 \perp e_1$, the maps in the rows of f_M are disjoint by the stability of the disjointness relation. Similarly, the maps in the columns are disjoint by universality. We have $e_0 f e_0 \sqcup e_0 f e_1 = e_0 f$ and $e_1 f e_0 \sqcup e_1 f e_1 = e_1 f$. Each of these maps are disjoint by universality. Finally, $e_0 f \sqcup e_1 f = (e_0 \sqcup e_1) f = f$ and hence we may recover the initial map whenever we have a matrix of this form. We will call this computation the distinct join of f_M .

Next, consider $f_M \times g_M$. As each e_i is idempotent, this is

$$f_M \times g_M = \begin{bmatrix} e_0 f e_0 g e_0 \sqcup e_0 f e_1 g e_0 & e_0 f e_0 g e_1 \sqcup e_0 f e_1 g e_1 \\ e_1 f e_0 g e_0 \sqcup e_1 f e_1 g e_0 & e_1 f e_0 g e_1 \sqcup e_1 f e_1 g e_1 \end{bmatrix} = \begin{bmatrix} e_0 f g e_0 & e_0 f g e_1 \\ e_1 f g e_0 & e_1 f g e_1 \end{bmatrix}$$

where the distinct joins are well defined due to the stability and universality of the join. We can see that the distinct join of $f_M \times g_M = f g$ and as such we have composition.

□

In particular, we note that we may represent $f : A \rightarrow B$ by the matrix

$$\begin{bmatrix} 1f1 & 1f0 \\ 0f1 & 0f0 \end{bmatrix} = \begin{bmatrix} f & 0 \\ 0 & 0 \end{bmatrix}$$

as $A \cong A \oplus 0$ and $B \cong B \oplus 0$.

We now turn to examining a category of specialized matrices over an inverse sum category. In general, the matrix category $\mathbf{iMat}(\mathbb{X})$ will have objects that are lists of objects in \mathbb{X} , $X = (X_1, \dots, X_m)$. Maps between lists will be matrices $[f_{ij}] : (X_1, \dots, X_m) \rightarrow (Y_1, \dots, Y_n)$. We will only consider maps whose matrices have disjoint rows, i.e., if $[f_{ij}]$ is a matrix, it must have $f_{ij} \perp f_{ik}$ for all i whenever $j \neq k$.

Lemma 3.6.22. *If \mathbb{X} is an inverse category with an inverse sum tensor, $\mathbf{iMat}(\mathbb{X})$, the category of inverse sum matrices over \mathbb{X} with composition as in Definition 3.6.18, has sums.*

Proof.

□

3.7 Completing a distributive inverse category

3.7.1 Distributive restriction categories

Definition 3.7.1. A Cartesian restriction category with a restriction zero and coproducts is called *distributive* when there is an isomorphism ρ such that

$$A \times (B + C) \xrightarrow{\rho} (A \times B) + (A \times C).$$

In a distributive inverse category, we lack:

$$A \xrightarrow{!} 1$$

and

$$A + A \xrightarrow{\nabla} A.$$

3.7.2 Distributive inverse categories

A *Distributive inverse category* has a sum tensor and product tensor as defined in sub-section [sub-section 3.7.1](#) where the product distributes over the sum in the following manner:

$$(A \otimes B) \oplus (A \otimes C) \xrightarrow{f} A \otimes (B \oplus C)$$

where f is some total function. As we are in an inverse category and f is total, we have $f f^{(-1)} = 1$. This also implies $1 = f^{(-1)} f$ where

$$A \otimes (B \oplus C) \xrightarrow{f^{(-1)}} (A \otimes B) \oplus (A \otimes C).$$

Suppose we have an inverse category \mathbb{X} with two tensors, \otimes and \oplus , as described above and we have:

$$\begin{array}{ccc} A \otimes (B \oplus C) & \xrightarrow{[1 \otimes \Pi_0, 1 \otimes \Pi_1]^{(-1)}} & (A \otimes B) \oplus (A \otimes C) \\ & \searrow & \downarrow [1 \otimes \Pi_0, 1 \otimes \Pi_1] \\ & & A \otimes (B \oplus C) \end{array}$$

Lemma 3.7.2. *The category \mathbb{X} as defined above is an inverse distributive category.*

Proof. We will show that the required function, f is $[1 \otimes \Pi_0, 1 \otimes \Pi_1]$.

$$(A \otimes B) \oplus (A \otimes C) \xrightarrow{[1 \otimes \Pi_0, 1 \otimes \Pi_1]} A \otimes (B \oplus C) \quad (3.20)$$

The map in [equation \(3.20\)](#) is total. To see this, consider the unique decomposition of the restriction of the map.

$$\begin{aligned} \overline{[1 \otimes \Pi_0, 1 \otimes \Pi_1]} &= \begin{bmatrix} \overline{1 \otimes \Pi_0} & 0 \\ 0 & \overline{1 \otimes \Pi_1} \end{bmatrix} \\ &= \begin{bmatrix} \bar{1} \otimes \overline{\Pi_0} & 0 \\ 0 & \bar{1} \otimes \overline{\Pi_1} \end{bmatrix} \\ &= \begin{bmatrix} 1 \otimes 1 & 0 \\ 0 & 1 \otimes 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= 1 \end{aligned}$$

□

Chapter 4

Quantum computation and circuits

4.1 Linear algebra

Quantum computation requires familiarity with the basics of linear algebra. This section will give definitions of the terms used throughout this thesis.

4.1.1 Basic definitions

The first definition needed is that of a *vector space*.

Definition 4.1.1 (Vector Space). Given a field F , whose elements will be referred to as scalars, a *vector space* over F is a non-empty set V with two operations, *vector addition* and *scalar multiplication*. *Vector addition* is defined as $+$: $V \times V \rightarrow V$ and denoted as $\mathbf{v} + \mathbf{w}$ where $\mathbf{v}, \mathbf{w} \in V$. The set V must be an abelian group under $+$. *Scalar multiplication* is defined as \cdot : $F \times V \rightarrow V$ and denoted as $c\mathbf{v}$ where $c \in F, \mathbf{v} \in V$. Scalar multiplication distributes over both vector addition and scalar addition and is associative. F 's multiplicative identity is an identity for scalar multiplication.

The specific algebraic requirements are:

1. $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V, (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w});$
2. $\forall \mathbf{u}, \mathbf{v} \in V, \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u};$
3. $\exists \mathbf{0} \in V$ such that $\forall \mathbf{v} \in V, \mathbf{0} + \mathbf{v} = \mathbf{v};$
4. $\forall \mathbf{u} \in V, \exists \mathbf{v} \in V$ such that $\mathbf{u} + \mathbf{v} = \mathbf{0};$
5. $\forall \mathbf{u}, \mathbf{v} \in V, c \in F, c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v};$

$$6. \forall \mathbf{u} \in V, c, d \in F, (c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u};$$

$$7. \forall \mathbf{u} \in V, c, d \in F, (cd)\mathbf{u} = c(d\mathbf{u});$$

$$8. \forall \mathbf{u} \in V, 1\mathbf{u} = \mathbf{u}.$$

Examples of vector spaces over F are: $F^{n \times m}$ – the set of $n \times m$ matrices over F ; and F^n – the n –fold Cartesian product of F . $F^{n \times 1}$, the set of $n \times 1$ matrices over F is also called the space of column vectors, while $F^{1 \times n}$, the set of row vectors. Often, F^n is identified with $F^{n \times 1}$.

This thesis shall identify F^n with the column vector space over F .

Definition 4.1.2 (Linearly independent). A subset of vectors $\{\mathbf{v}_i\}$ of the vector space V is said to be *linearly independent* when no finite linear combination of them, $\sum a_j \mathbf{v}_j$ equals $\mathbf{0}$ unless all the a_j are zero.

Definition 4.1.3 (Basis). A *basis* of a vector space V is a linearly independent subset of V that generates V . That is, any vector $u \in V$ is a linear combination of the basis vectors.

4.1.2 Matrices

As mentioned above, the set of $n \times m$ matrices over a field is a vector space. Additionally, matrices compose and the tensor product of matrices is defined.

Matrix composition is defined as usual. That is, for $A = [a_{ij}] \in F^{m \times n}$, $B = [b_{jk}] \in F^{n \times p}$:

$$AB = \left[\left(\sum_j a_{ij} b_{jk} \right)_{ik} \right] \in F^{m \times p}.$$

Definition 4.1.4 (Diagonal matrix). A *diagonal matrix* is a matrix where the only non-zero entries are those where the column index equals the row index.

The diagonal matrix $n \times n$ with only 1's on the diagonal is the identity for matrix multiplication, and is designated by I_n .

Definition 4.1.5 (Transpose). The *transpose* of an $n \times m$ matrix $A = [a_{ij}]$ is an $m \times n$ matrix A^t with the i, j entry being a_{ji} .

When the base field of a matrix is \mathbb{C} , the complex numbers, the *conjugate transpose* (also called the *adjoint*) of an $n \times m$ matrix $A = [a_{ij}]$ is defined as the $m \times n$ matrix A^* with the i, j entry being \bar{a}_{ji} , where \bar{a} is the complex conjugate of $a \in \mathbb{C}$.

When working with column vectors over \mathbb{C} , note that $\mathbf{u} \in \mathbb{C}^n \implies \mathbf{u}^* \in \mathbb{C}^{1 \times n}$ and that $\mathbf{u}^* \times \mathbf{u} \in \mathbb{C}^{1 \times 1}$. This thesis will use the usual identification of \mathbb{C} with $\mathbb{C}^{1 \times 1}$. A column vector \mathbf{u} is called a *unit vector* when $\mathbf{u}^* \times \mathbf{u} = 1$.

Definition 4.1.6 (Trace). The *trace*, $Tr(A)$ of a square matrix $A = [a_{ij}]$ is $\sum a_{ii}$.

Tensor Product

The tensor product of two matrices is the usual Kronecker product:

$$U \otimes V = \begin{bmatrix} u_{11}V & u_{12}V & \cdots & u_{1m}V \\ u_{21}V & u_{22}V & \cdots & u_{2m}V \\ \vdots & \vdots & \ddots & \\ u_{n1}V & u_{n2}V & \cdots & u_{nm}V \end{bmatrix} = \begin{bmatrix} u_{11}v_{11} & \cdots & u_{12}v_{11} & \cdots & u_{1m}v_{1q} \\ u_{11}v_{21} & \cdots & u_{12}v_{21} & \cdots & u_{1m}v_{2q} \\ \vdots & \vdots & \vdots & \ddots & \\ u_{n1}v_{p1} & \cdots & u_{n2}v_{p1} & \cdots & u_{nm}v_{pq} \end{bmatrix}$$

Special matrices

When working with quantum values certain types of matrices over the complex numbers are of special interest. These are:

Unitary Matrix : Any $n \times n$ matrix A with $AA^* = I$ ($= A^*A$).

Hermitian Matrix : Any $n \times n$ matrix A with $A = A^*$.

Positive Matrix : Any Hermitian matrix A in $\mathbb{C}^{n \times n}$ where $\mathbf{u}^* A \mathbf{u} \geq 0$ for all vectors $\mathbf{u} \in \mathbb{C}^n$.

Note that for any Hermitian matrix A and vector u , $\mathbf{u}^* A \mathbf{u}$ is real.

Completely Positive Matrix : Any positive matrix A in $\mathbb{C}^{n \times n}$ where $I_m \otimes A$ is positive.

The matrix

$$\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

is an example of a matrix that is *unitary*, *Hermitian*, *positive* and *completely positive*.

Superoperators

A *Superoperator* S is a matrix over \mathbb{C} with the following restrictions:

1. S is *completely positive*. This implies that S is positive as well.
2. For all positive matrices A , $Tr(S A) \leq Tr(A)$.

4.2 Basic quantum computation

4.2.1 Quantum bits

Quantum computation deals with operations on **qubits**. A **qubit** is typically represented in the literature on quantum computation as a complex linear combination of $|0\rangle$ and $|1\rangle$, respectively identified with $(1,0)$ and $(0,1)$ in \mathbb{C}^2 . Because of the identification of the basis vectors, any **qubit** can be identified with a non-zero vector in \mathbb{C}^2 . In standard quantum computation, the important piece of information in a **qubit** is its direction rather than amplitude. In other words, given $q = \alpha |0\rangle + \beta |1\rangle$ and $q' = \alpha' |0\rangle + \beta' |1\rangle$ where $\alpha = \gamma\alpha'$ and $\beta = \gamma\beta'$, then q and q' represent the same quantum state.

A **qubit** that has either α or β zero is said to be in a *classical state*. Any other combination of values is said to be a *superposition*.

[Section 4.3 on page 130](#) will introduce quantum circuits which act on **qubits**. This section will have some forward references to circuits to illustrate points introduced here.

4.2.2 Quantum entanglement

Consider what happens when working with a pair of **qubits**, p and q . This can be considered as the a vector in \mathbb{C}^4 and written as

$$\alpha_{00} |00\rangle + \alpha_{01} |01\rangle + \alpha_{10} |10\rangle + \alpha_{11} |11\rangle. \quad (4.1)$$

In the case where p and q are two independent **qubits**, with $p = \alpha |0\rangle + \beta |1\rangle$ and $q = \gamma |0\rangle + \delta |1\rangle$,

$$p \otimes q = \alpha\gamma |00\rangle + \alpha\delta |01\rangle + \beta\gamma |10\rangle + \beta\delta |11\rangle \quad (4.2)$$

where $p \otimes q$ is the standard tensor product of p and q regarded as vectors. There are states of two **qubits** that cannot be written as a tensor product. As an example, the state

$$\frac{1}{\sqrt{2}} |00\rangle + \frac{1}{\sqrt{2}} |11\rangle \quad (4.3)$$

is not a tensor product of two distinct **qubits**. In this case the two **qubits** are said to be *entangled*.

4.2.3 Quantum gates

Quantum gates operate on **qubits**. These gates are conceptually similar to logic gates in the classical world. In the classical world the only non-trivial single **bit** gate is the Not gate which sends 0 to 1 and 1 to 0. However, there are infinitely many non-trivial quantum gates.

An n -**qubit** quantum gate is represented by a $2^n \times 2^n$ matrix. A necessary and sufficient condition for such a matrix to be a quantum gate is that it is *unitary*.

The entanglement of two **qubits**, p and q , is accomplished by applying a Hadamard transformation to p followed by a Not applied to q controlled by p . The circuit in [figure 4.2 on page 131](#) shows how to entangle two **qubits** that start with an initial state of $|00\rangle$. See this can be done in L-QPL.

A list of some common gates, together with their usual quantum circuit representation is given in the next section in [table 4.1 on page 132](#).

4.2.4 Measurement

The other allowed operation on a **qubit** or group of **qubits** is measurement. When a **qubit** is measured it assumes only one of two possible values, either $|0\rangle$ or $|1\rangle$. Given

$$q = \alpha |0\rangle + \beta |1\rangle \quad (4.4)$$

where $|\alpha|^2 + |\beta|^2 = 1$, then measuring q will result in $|0\rangle$ with probability $|\alpha|^2$ and $|1\rangle$ with probability $|\beta|^2$. Once a **qubit** is measured, re-measuring will always produce the same value.

In multi-**qubit** systems the order of measurement does not matter. If p and q are as in [equation \(4.1\) on the previous page](#), let us suppose measuring p gives $|0\rangle$. The measure will result in that value with probability $|\alpha_{00}|^2 + |\alpha_{01}|^2$, after which the system collapses to the state:

$$\alpha_{00} |00\rangle + \alpha_{01} |01\rangle \quad (4.5)$$

Measuring the second **qubit**, q , will give $|0\rangle$ with probability $|\alpha_{00}|^2$ or $|1\rangle$ with probability $|\alpha_{01}|^2$.

Conversely, if q was measured first and gave us $|0\rangle$ (with a probability of $|\alpha_{00}|^2 + |\alpha_{10}|^2$) and then p was measured, p will give us $|0\rangle$ with probability $|\alpha_{00}|^2$ or $|1\rangle$ with probability $|\alpha_{10}|^2$.

Thus, when measuring both p and q , the probability of getting $|0\rangle$ from both measures is $|\alpha_{00}|^2$, regardless of which **qubit** is measured first.

Considering states such as in [equation \(4.3\) on the preceding page](#), measuring either **qubit** would actually force the other **qubit** to the same value. This type of entanglement is used in many quantum algorithms such as quantum teleportation.

4.2.5 Mixed states

The notion of *mixed states* refers to an outside observer's knowledge of the state of a quantum system. Consider a 1 **qubit** system

$$\nu = \alpha |0\rangle + \beta |1\rangle. \quad (4.6)$$

If ν is measured but the results of the measurement are not examined, the state of the system is either $|0\rangle$ or $|1\rangle$ and is no longer in a superposition. This type of state is written as:

$$\nu = |\alpha|^2 \{|0\rangle\} + |\beta|^2 \{|1\rangle\}. \quad (4.7)$$

An external (to the state) observer knows that the state of ν is as expressed in **equation (4.7)**. Since the results of the measurement were not examined, the exact state (0 or 1) is unknown. Instead, a probability is assigned as expressed in the equation. Thus, if the **qubit** ν is measured and the results are not examined, ν can be treated as a probabilistic **bit** rather than a **qubit**.

4.2.6 Density matrix notation

The state of any quantum system of **qubits** may be represented via a *density matrix*. In this notation, given a **qubit** ν , the coefficients of $|0\rangle$ and $|1\rangle$ form a column vector u . Then the density matrix corresponding to ν is uu^* . If $\nu = \alpha |0\rangle + \beta |1\rangle$,

$$\nu = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{pmatrix} \bar{\alpha} & \bar{\beta} \end{pmatrix} = \begin{pmatrix} \alpha\bar{\alpha} & \alpha\bar{\beta} \\ \beta\bar{\alpha} & \beta\bar{\beta} \end{pmatrix}. \quad (4.8)$$

When working with mixed states the density matrix of each component of the mixed state is added. For example, the mixed state shown in **equation (4.7)** would be represented by the density matrix

$$|\alpha|^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + |\beta|^2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix}. \quad (4.9)$$

Note that since the density matrix of mixed states is a linear combination of other density matrices, it is possible to have two different mixed states represented by the same density matrix.

The advantage of this notation is that it becomes much more compact for mixed state systems. Additionally, scaling issues are handled by insisting the density matrix has a trace $= 1$. During a general quantum computation, as we shall see, the trace can actually fall below 1 indicating that the computation is not everywhere total.

4.2.7 Gates and density matrices

When considering a **qubit** q as a column vector and a unitary transform T as a matrix, the result of applying the transform T to q is the new vector Tq . The density matrix of the original **qubit** is given by qq^* , while the density matrix of the transformed **qubit** is $(Tq)(Tq)^*$, which equals $T(qq^*)T^*$. Thus, when a **qubit** q is represented by a density matrix A , the formula for applying the transform T to q is TAT^* .

4.3 Quantum circuits

4.3.1 Contents of quantum circuits

Currently a majority of quantum algorithms are defined and documented using *quantum circuits*. These are wire-type diagrams with a series of **qubits** input on the left of the diagram and output on the right. Various graphical elements are used to describe quantum gates, measurement, control and classical **bits**.

Gates and **qubits**

The simplest circuit is a single wire with no action:

$$\text{---} \underline{x} \text{---}$$

The next simplest circuit is one **qubit** and one gate. The **qubit** is represented by a single wire, while the gate is represented by a box with its name, G , inside it. This is shown in the circuit in [figure 4.1](#). In general, the name of the wire which is input to the gate G may be different from the name of G 's output wire. Circuit diagrams may also contain constant components as input to gates as in the circuit in [figure 4.3](#).

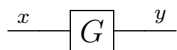


Figure 4.1: Simple single gate circuit

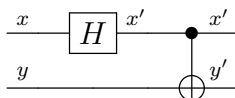


Figure 4.2: Entangling two **qubits**.

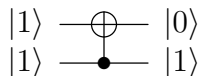


Figure 4.3: Controlled-Not of $|1\rangle$ and $|1\rangle$

Future diagrams will drop the wire labels except when they are important to the concept under discussion.

Controlled gates, where the gate action depends upon another **qubit**, are shown by attaching a wire between the wire of the control **qubit** and the controlled gate. The circuit in [figure 4.2](#) shows two **qubits**, where a Hadamard is applied to the top **qubit**, followed by a Controlled-Not applied to the second **qubit**. In circuits, the control **qubit** is on the vertical wire with the solid dot. This is then connected via a horizontal wire to the gate being controlled.

A list of common gates, their circuits and corresponding matrices is given in [table 4.1 on the following page](#).

Measurement

Measurement is used to transform the quantum data to classical data so that it may be then used in classical computing (e.g. for output). The act of measurement is placed at the last







Gate	Circuit	Matrix
Not (X)		$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
Z		$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Hadamard		$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$
Swap		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Controlled-Not		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$
Toffoli		$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

Table 4.1: Gates, circuit notation and matrices

part of the quantum algorithm in many circuit diagrams and is sometimes just implicitly considered to be there.

While there are multiple notations used for measurement in quantum circuit diagrams, this thesis will standardize on the *D-box* style of measurement as shown in [figure 4.4](#).

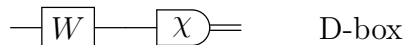


Figure 4.4: Measure notation in quantum circuits

A measurement may have a double line leaving it, signifying a **bit**, or nothing, signifying a destructive measurement.

Operations affecting multiple **qubits** at the same time are shown by extending the gate or measure box to encompass all desired wires. In the circuit in [figure 4.5](#), the gate U applies to all of the first three **qubits** and the measurement applies to the first two **qubits**.

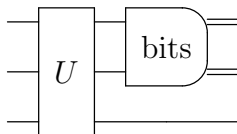


Figure 4.5: Examples of multi-**qubit** gates and measures

0-control and control by **bits**

The examples above have only shown control based upon a **qubit** being $|1\rangle$. Circuits also allow control on a **qubit** being $|0\rangle$ and upon classical values. These forms of control are illustrated by the circuit in [figure 4.6 on the next page](#) with four **qubits** (r_1, r_2, p and q).

At g_1 , a Hadamard is 1-controlled by r_2 and is applied to each of r_1 and p . This is followed in column g_2 with the Not transform applied to r_2 being 0-controlled by r_1 . In the same column, a Z gate is 0-controlled by q and applied to p . p and q are then measured in column g_3 and their corresponding classical values are used for control in g_4 . In g_4 , the U_R gate is applied to both r_1 and r_2 , but only when the measure result of p is 0 and the measure result of q is 1.

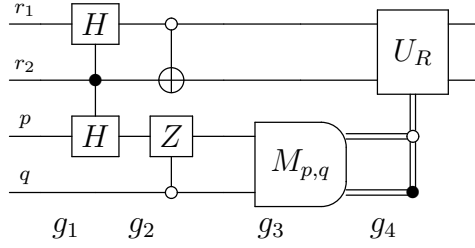


Figure 4.6: Other forms of control for gates

Multi-qubit lines

It is common to represent multiple **qubits** on one line. A gate applied to a multi-**qubit** line must be a tensor product of gates of the correct dimensions. The circuit in [figure 4.7](#) shows n **qubits** on one line with the Hadamard gate (tensored with itself n times) applied to all of them. That is followed by a unary gate U_R tensored with $I^{\otimes(n-2)}$ and tensored with itself again. This will have the effect of applying an U_R gate to the first and last **qubits** on the line.

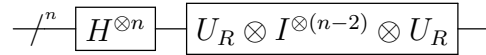


Figure 4.7: n **qubits** on one line

Other common circuit symbols

Two other symbols that are regularly used are the swap and controlled- Z , shown in the circuit in [figure 4.8](#). Note that swap is just shorthand for a series of three controlled-Not gates with the control **qubit** changing. This can also be seen directly by multiplying the matrices for the controlled-Not gates as shown in [equation \(4.10\) on the following page](#).

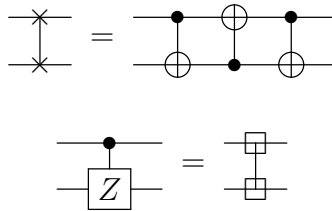


Figure 4.8: Swap and controlled- Z

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (4.10)$$

4.3.2 Syntax of quantum circuits

Quantum circuits were originally introduced by David Deutsch in [10]. He extended the idea of standard classical based gate diagrams to encompass the quantum cases. In his paper, he introduced the concepts of quantum gates, sources of **bits**, sinks and universal gates. One interesting point of the original definition is that it *does* allow loops. Currently, the general practice is not to allow loops of **qubits**. The commonly used elements of a circuit are summarized in [table 4.2 on the next page](#).

A valid quantum circuit must follow certain restrictions. As physics requires **qubits** must not be duplicated, circuits must enforce this rule. Therefore, three restrictions in circuits are the *no fan-out*, *no fan-in* and *no loops* rules. These conditions are a way to express the *linearity* of quantum algorithms. Variables (wires) may not be duplicated, may not be destroyed without a specific operation and may not be amalgamated.

4.3.3 Examples of quantum circuits

This section will present three quantum algorithms and the associated circuits. Each of these circuits presented may be found in [18].

First, *quantum teleportation*, an algorithm which sends a quantum bit across a distance via the exchange of two classical bits. This is followed by the *Deutsch-Jozsa algorithm*, which provides information about the global nature of a function with less work than a classical deterministic algorithm can. The third example is circuits for the *quantum Fourier transformation* and its inverse.


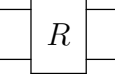
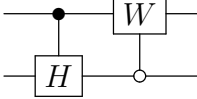

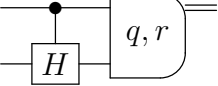
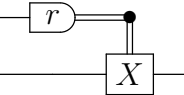
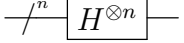
Desired element	Element in a quantum circuit diagram.	Example
qubit	A single horizontal line.	—
Classical bit	A double horizontal line.	=
Single- qubit gates	A box with the gate name (G) inside it, one wire attached on its left and one wire attached on the right.	
Multi- qubit gates	A box with the gate name (R) inside it, n wires on the left side and the same number of wires on the right.	
Controlled qubit gates	A box with the gate name (H, W) inside, with a solid (1-control) or open (0-control) dot on the control wire with a vertical wire between the dot and the second gate.	
Controlled-Not gates	A <i>target</i> \oplus , with a solid (1-control) or open (0-control) dot on the control wire with a vertical wire between the dot and the gate.	
Measurement	A D shaped node with optional names or comments inside. One to n single wires are attached on the left (qubits coming in) and 0 to m classical bit wires on the right where $m \leq n$. Classical bits may be dropped as desired.	
Classical control	Control bullets are attached to horizontal classical wires, with vertical classical wires attached to the controlled gate.	
Multiple qubits	Annotate the line with the number of qubits and use tensors on gates.	

Table 4.2: Syntactic elements of quantum circuit diagrams

Quantum teleportation

The standard presentation of this algorithm involves two participants A and B . (Henceforth known as Alice and Bob). Alice and Bob first initialize two **qubits** to $|00\rangle$, then place them into what is known as an *EPR* (for Einstein, Podolsky and Rosen) state. This is accomplished by first applying the Hadamard gate to Alice's **qubit**, followed by a Controlled-Not to the pair of **qubits** controlled by Alice's **qubit**.

Then, Bob travels somewhere distant from Alice, taking his **qubit** with him¹.

At this point, Alice receives a **qubit**, ν , in an unknown state and has to pass ν on to Bob. She then uses ν as the control and applies a Controlled-Not transform to this new pair. Alice then applies a Hadamard transform applied to ν .

Alice now measures the two **qubits** and sends the resulting two **bits** of classical information to Bob.

Bob then examines the two **bits** that he receives from Alice. If the **bit** resulting from measuring Alice's original bit is 1, he applies the Not (also referred to as X) gate ($= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$) to his **qubit**. If the measurement result of ν is one, he applies the Z gate ($= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$). Bob's **qubit** is now in the same state as the **qubit** Alice wanted to send. The circuit for this is shown in [figure 4.9](#).

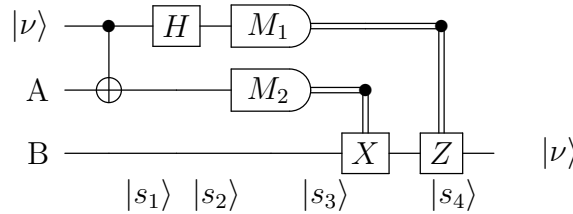


Figure 4.9: Quantum teleportation

For comparison , see showing how this would be implemented in L-QPL.

¹Notice that all other physical constraints are ignored in this algorithm. There is no concern about how one separates the **qubits**, transports the **qubit** or potential decoherence of the **qubits**.

Deutsch-Jozsa algorithm

The Deutsch-Jozsa algorithm describes a way of determining whether a function f^2 is *constant* (i.e. always 0 or 1) or *balanced* (i.e. produces an equal number of 0 or 1 results) based on applying it to one quantum bit. The function takes n **bits** as input and produces a single **bit**.

f is assumed to be an expensive function, therefore, a desired effect is to evaluate f as few times as possible before determining if f is balanced or constant. The worst case scenario when evaluating f classically is that determining the result requires $2^{n-1} + 1$ invocations of the function. The best possible case is 2 invocations, which occurs when f is balanced and the first two inputs chosen produce different results.

The quantum circuit requires only one application of the function to $n + 1$ **qubits** which have been suitably prepared to make the decision.

The algorithm relies on being able to construct an $n + 1$ order unitary operator based upon f . In general, a unitary operator like this may be constructed by mapping the state $|a, b\rangle$ to $|a, b \oplus f(a)\rangle$ where \oplus is the exclusive-or operator and a is n **bit** values. If we name this operator U_f , the circuit in [figure 4.10](#) will solve the problem with just one application. See the appendix, for how this would be done in L-QPL.

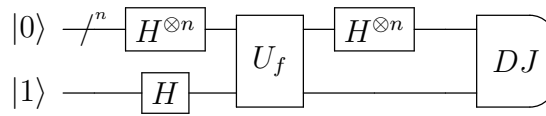


Figure 4.10: Circuit for the Deutsch-Jozsa algorithm

The idea of quantum parallelism is what makes this and many other quantum algorithms work. The initial state of the system is set to $|0^{\otimes n} \otimes 1\rangle$ after which the Hadamard gate is applied to all of the **qubits**. This places the input **qubits** into a superposition of all possible input values and the answer **qubit** is a superposition of 0 and 1. At this point, the unitary

²The obvious pre-condition for the Deutsch-Jozsa algorithm is that the function f is *either* balanced or constant and not some general function. The results are not well-defined if f does not fit into one of the two possible categories.

transformation U_f is applied to the **qubits**. Then the Hadamard transform is applied again to the input **qubits**.

To complete the algorithm, measure *all* the **qubits**. It can be shown that if f is constant, the input **qubits** will all measure to 0, while if it is balanced, at least one of those **qubits** will be 1.

Quantum Fourier transform

The circuits for the quantum Fourier transformation and its inverse are in [figure 4.11](#) and [figure 4.12 on the next page](#) respectively. These transforms are used extensively in many quantum algorithms, including Shor's factoring algorithm.

The quantum Fourier transform is definable on an arbitrary number of **qubits**. This is typically presented by eliding the 3rd to the $n - 3^{\text{rd}}$ lines and interior processing. The L-QPL code for the quantum Fourier transform is in the appendix, In this circuit, the parametrized transform R_n is the rotation transform, given by:

$$R_n = \begin{bmatrix} 1 & 0 \\ 0 & e^{\frac{2\pi i}{2^n}} \end{bmatrix}$$

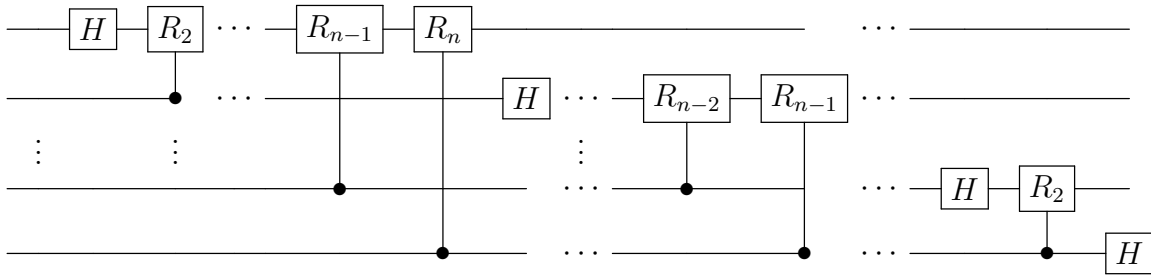


Figure 4.11: Circuit for the quantum Fourier transform

The inverse of a circuit is determined by traversing the circuit from right to left. This process changes the original quantum Fourier circuit to its inverse as shown in [figure 4.12 on the following page](#). The L-QPL code for the inverse quantum Fourier transform is in the appendix,

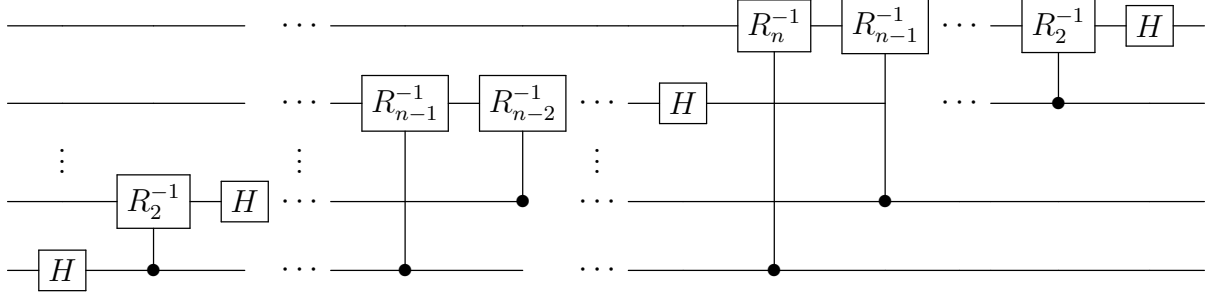


Figure 4.12: Circuit for the inverse quantum Fourier transform

4.4 Extensions to quantum circuits

To facilitate the transition to the programming language L-QPL, this section introduces three extensions to quantum circuits. The extensions are *renaming*, *wire bending and crossing*, and *scoped control*. Each extension adds expressive power to quantum circuits but does not change the semantic power. For each of the extensions, examples of how to re-write the extension in standard quantum circuit terminology will be provided.

4.4.1 Renaming

Quantum circuits currently allow renaming to be an implicit part of any gate. The circuit in [figure 4.13](#) gives an operation to explicitly do this and its rewriting in standard circuit notation.

$$\frac{y}{\text{---}} \boxed{x := y} \frac{x}{\text{---}} \equiv \frac{y}{\text{---}} \boxed{I} \frac{x}{\text{---}}$$

Figure 4.13: Renaming of a **qubit** and its equivalent diagram

4.4.2 Wire crossing

Crossing and bending of wires in a circuit diagram is added to allow a simpler presentation of algorithms. The circuit in [figure 4.14 on the following page](#) illustrates the concept of re-organizing and bending of wires.

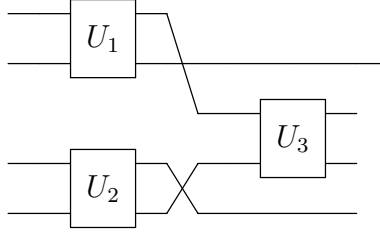


Figure 4.14: Bending

4.4.3 Scoped control

This extension allows us to group different operations in a circuit and show that all of them are controlled by a particular **qubit**. This is the same as attaching separate control wires to each of the gates in the grouped operations. Measurements are not affected by control. **Figure 4.15** shows a scoped control box on the left which includes a measure. The right hand side of the same figure shows the circuit translated back to standard circuit notation, with the measure not being affected by the control.

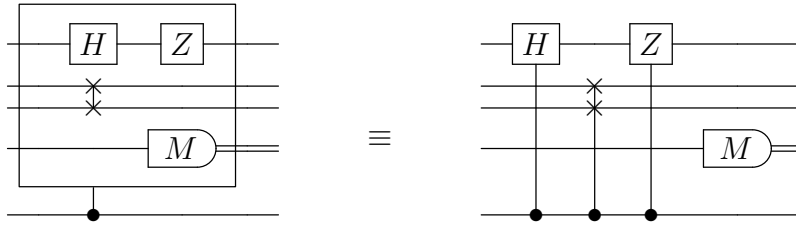


Figure 4.15: Scope of control

Scoping boxes correspond to procedures and blocks in L-QPL.

Naturally, both scoping and bending may be combined as in **figure 4.16**.

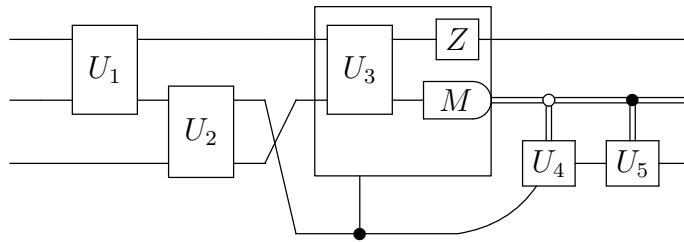


Figure 4.16: Extensions sample

However, note that exchanging wires is not the same as swap. Exchanging a pair of wires is not affected by control, but a swap is affected by control, as shown in [figure 4.17](#).

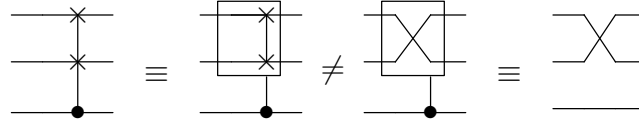


Figure 4.17: Swap in control vs. exchange in control

4.4.4 Circuit identities

Circuits allow the writing of the same algorithm in multiple ways. This sub-section will list some of the circuit identities that hold with the extended notation.

First, note that although a measure may appear inside a control box, it is not affected by the control, as in [figure 4.18](#). Conversely, a measurement commutes with control of a circuit as in [figure 4.18](#).

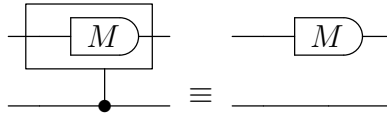


Figure 4.18: Measure is not affected by control

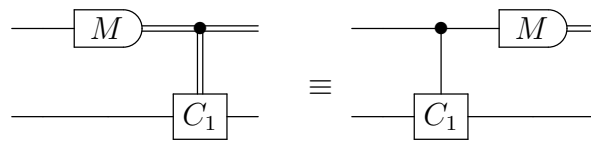


Figure 4.19: Control is not affected by measure

One of the notations introduced earlier was that of *0-control*. This type of control is the same as applying a *Not* transform before and after a *1-control*, as shown in [figure 4.20 on the following page](#).

[Figure 4.21 on the next page](#) shows that scoped control of multiple transforms is the same as controlling those transforms individually. [Figure 4.22 on the following page](#) similarly

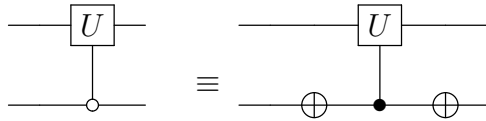


Figure 4.20: Zero control is syntactic sugar

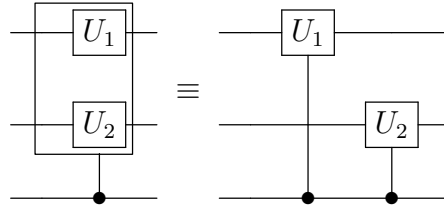


Figure 4.21: Scoped control is parallel control

shows that scoped control of multiple transforms of the same **qubit** is the same as controlling those transforms serially.

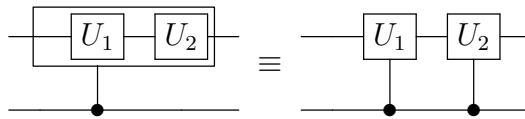


Figure 4.22: Scoped control is serial control

Multiple control commutes with scoping as shown in [figure 4.23](#) to [figure 4.24](#) on the next page.

4.5 An alternate description of quantum circuits

In order to explore transforms of quantum circuits, it is helpful to have a algebraic description which will allow manipulation of the circuits. The goals of the algebraic description are:

- Represent **qubits** and **bits**;
- represent gates;
- allow algorithmic manipulations of the circuit;
- allow proving that correctness of manipulations.

Note that measurement is not included in this representation.

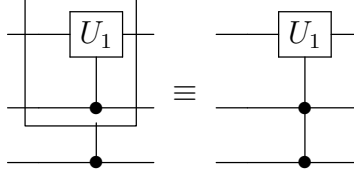


Figure 4.23: Multiple control

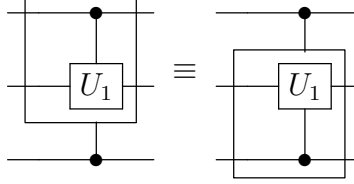


Figure 4.24: Control scopes commute

4.5.1 Base types

The base types **bit** and **qubit** are needed for any circuit. These will be taken as elemental and correspond to the classical notions of “bit” (i.e., 0 or 1) and “qubit” (i.e., $\alpha|0\rangle + \beta|1\rangle$).

Definition 4.5.1. The type **endpoint** is either **bit** or **qubit**.

Definition 4.5.2. The type **Q – Arity** is a partial map from $W \subset \mathbb{N}$ to **Endpoint**.

Definition 4.5.3. A *wire* is an element of $W \subset \mathbb{N}$. A *typed wire* is a wire together with a specified **Q – Arity**.

Definition 4.5.4. A *control wire* is a pair (w, b) where w is a typed wire and $b :: \mathbf{Bool}$.

A quantum program uses *typed wires* as its data.

Definition 4.5.5. A *gate* is a function from W_1 , a set of typed wires, to W_2 , another set of typed wires. The gate function must be a superoperator as defined in [23].

4.5.2 Types and Shapes

Although circuits, gates and low level subroutines are defined at the level of wires, programmatically, we would like to refer to groupings of wires. For example, a list of **qubit** or a register (tuple) of **qubit**. These groupings may have a required leaf type (e.g., **qubit** or

bit) or it may be considered polymorphic. Additionally, the leaf type may be mixed or homogenous.

Definition 4.5.6. The singleton type **B** is defined as having the instance $B_{()}$ and the singleton type **Q** is defined as having the singleton instance $Q_{()}$.

Definition 4.5.7. The type family **QCData** consists of algebraic data types built up from **bit** or **qubit**.

Examples of **QCData** include $(\mathbf{bit}, \mathbf{qubit}, [\mathbf{qubit}])$ and **bit**.

Definition 4.5.8. Given an instance I of a type **T** in **QCData**, the *shape* of I is the instance I_s obtained from I by replacing all terms of type **bit** by $B_{()}$ and all terms of type **qubit** by $Q_{()}$.

For example, a 2 element list of **qubits** has shape $[Q_{()}, Q_{()}]$ while a pair of a **qubit** and **bit** has shape $(Q_{()}, B_{()})$

The concepts of shape and the type **QCData** allow us to group the wires of a quantum circuit into higher order types.

4.6 Quantum computation

Quantum computation proceeds via the application of reversible transformations — Unitary transformations.

The semantics of quantum computation can be defined as a \dagger -compact closed category as introduced in [2, 3] and completely positive maps as discussed in [24].

Definition 4.6.1 (Dagger Category). A *Dagger Category* [24] is a category \mathbb{C} together with an operation \dagger that is an involutive, identity on objects, contra-variant endofunctor on \mathbb{C} .

Recalling first that a *symmetric monoidal category* is a category \mathbb{B} with a bi-functor \otimes , an object I and natural isomorphisms:

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

$$c_{A,B} : A \otimes B \rightarrow B \otimes A$$

$$ul_A : A \rightarrow I \otimes A$$

with standard coherence conditions, as in [15]. Note that we also have a map $ur_A : A \rightarrow A \otimes I$ given by $ur_A = ul_{Ac_{I,A}}$. Furthermore, a *compact closed category* \mathbb{C} is a symmetric monoidal category where each object A has a dual A^* together with the maps:

$$\eta_A : I \rightarrow A^* \otimes A$$

$$\epsilon_A : A \otimes A^* \rightarrow I$$

such that

$$\begin{array}{ccc} A \xrightarrow{ur_A} A \otimes I \xrightarrow{A \otimes \eta_A} A \otimes (A^* \otimes A) & \text{and} & A^* \xrightarrow{ul_{A^*}} I \otimes A^* \xrightarrow{\eta_{A^*} \otimes A^*} (A^* \otimes A) \otimes A^* \\ \parallel & & \parallel \\ & \downarrow a^{-1} & \downarrow a \\ & (A \otimes A^*) \otimes A & A^* \otimes (A \otimes A^*) \\ & \downarrow \epsilon \otimes A & \downarrow A^* \otimes \epsilon \\ & I \otimes A & A^* \otimes I \\ & \downarrow ul^{-1} & \downarrow ur^{-1} \\ & A & A \end{array}$$

From the above, we can define a *Dagger symmetric monoidal category* and a *Dagger compact closed category*. The latter is referred to as a *strongly compact closed category* in [2], where they were initially introduced. In each case, the \dagger functor is added in a way that retains coherence with the bi-functor \otimes and with the dualizing operator. The coherence implies that the $i^\dagger = i^{-1}$ for the SMC isomorphisms, that $(f \otimes g)^\dagger = f^\dagger \otimes g^\dagger$ for all maps f, g in the symmetric monoidal category and that

$$\begin{array}{ccc} I & \xrightarrow{\epsilon_A^\dagger} & A \otimes A^* \\ & \searrow \eta_A & \downarrow c \\ & & A^* \otimes A \end{array}$$

commutes for all objects A in the compact closed category.

Example 4.6.2 (REL). REL is a dagger compact closed category with the dual of an object A is A , \otimes is the cartesian product and for $R : A \rightarrow B$, we have $R^* = R^\dagger = \{(y, x) | (x, y) \in R\}$.

Example 4.6.3 (FDHILB). The category of finite dimensional Hilbert spaces, FDHILB is a dagger compact closed category with the dual of an object H is the normal Hilbert space dual H^* , the space of continuous linear functions from H to the base field. \otimes is the normal Hilbert space tensor and for $f : A \rightarrow B$, we have f^\dagger is the unique map such that $\langle fx | y \rangle = \langle y | f^\dagger x \rangle$ for all $x \in A, y \in B$.

Additionally, if one has a dagger compact closed category with biproducts where the biproducts and dagger interact such that $p_i^\dagger = q_i$, this is called a *biproduct dagger compact closed category*.

In [24], the author continues from this point: Starting with a biproduct dagger compact closed category \mathbb{C} , he creates a new category, $\text{CPM}(\mathbb{C})$ which has the same objects as \mathbb{C} , but morphisms $f : A \rightarrow B$ in $\text{CPM}(\mathbb{C})$ are given by maps $f : A^* \otimes A \rightarrow B^* \otimes B$ in \mathbb{C} which are *completely positive*. Note that REL and FDHILB are biproduct dagger compact closed categories.

From this, the category $\text{CPM}(\mathbb{C})^\oplus$, the free biproduct completion of $\text{CPM}(\mathbb{C})$ is formed, which is suitable for describing quantum computation semantics. For example, given FDHILB as our starting point, the tensor unit I is the field of complex numbers. The type of **qubit** (in FDHILB and by lifting, in $\text{CPM}(\text{FDHILB})^\oplus$) is given as $I \oplus I$. At this stage, the necessity of the CPM construction to model physical reality can be seen in the following as in FDHILB, the morphisms initialization of a qubit: $init : I \oplus I \rightarrow \mathbf{qubit}$ and destructive measure: $meas : \mathbf{qubit} \rightarrow I \oplus I$ are inverses. However, in $\text{CPM}(\text{FDHILB})^\oplus$, these same maps are given as

$$\mathbf{qubit}^* \otimes \mathbf{qubit} \xrightarrow{meas} I \oplus I \xrightarrow{init} \mathbf{qubit}^* \otimes \mathbf{qubit}$$

by the formulae:

$$meas \begin{pmatrix} a & b \\ c & d \end{pmatrix} = (a, d), \quad init(a, d) = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}.$$

Therefore, the maps are not inverses and reflect the physical reality.

Example 4.6.4 (Commutative Frobenius algebras). Let \mathbb{X} be a symmetric monoidal category and form $\text{CFrob}(\mathbb{X})$ as follows:

Objects: Commutative Frobenius algebras[14]: A quintuple $(X, \nabla, \eta, \Delta, \epsilon)$ where X is a k -algebra for some field k , and $\nabla : A \otimes A \rightarrow A$, $\eta : k \rightarrow A$, $\Delta : A \rightarrow A \otimes A$, $\epsilon : A \rightarrow k$ are natural maps in the algebra. Additionally, these satisfy

$$\begin{array}{ccc} A \otimes A & \xrightarrow{\Delta \otimes 1} & A \otimes (A \otimes A) \\ \downarrow 1 \otimes \Delta & \searrow \nabla & \downarrow 1 \otimes \nabla \\ (A \otimes A) \otimes A & \xrightarrow{\nabla \otimes 1} & A \otimes A \end{array}$$

together with the additional property that $\Delta \nabla = 1$.

Maps: Multiplication (∇) and co-multiplication (Δ) preserving homomorphisms which do not necessarily preserve the unit.

Theorem 4.6.5. *When \mathbb{X} is a symmetric monoidal category, $\text{CFrob}(\mathbb{X})$ is a discrete inverse category.*

Proof. For $f : X \rightarrow Y$, define $f^{(-1)}$ as

$$Y \xrightarrow{1 \otimes \eta} Y \otimes X \xrightarrow{1 \otimes \Delta} Y \otimes X \otimes X \xrightarrow{1 \otimes f \otimes 1} Y \otimes Y \otimes X \xrightarrow{\nabla \otimes 1} Y \otimes X \xrightarrow{\epsilon \otimes 1} X$$

Using a result from [20], we need only show:

$$(f^{(-1)})^{(-1)} = f$$

$$f f^{(-1)} f = f$$

$$f f^{(-1)} g g^{(-1)} = g g^{(-1)} f f^{(-1)}$$

We also use the following two identities from [14]:

$$(1 \otimes \eta)\nabla = id \quad (4.11)$$

$$\Delta(1 \otimes \epsilon) = id. \quad (4.12)$$

$$\begin{aligned}
f^{(-1)^{(-1)}} &= (1 \otimes \eta)(1 \otimes \Delta)(1 \otimes (f^{(-1)} \otimes 1)(\nabla \otimes 1)(\epsilon \otimes 1)) \\
&= (1 \otimes \eta)(1 \otimes \Delta)(1 \otimes ((1 \otimes \eta)(1 \otimes \Delta)(1 \otimes f \otimes 1)(\nabla \otimes 1)(\epsilon \otimes 1)) \otimes 1) \\
&\quad (\nabla \otimes 1)(\epsilon \otimes 1) \\
&= (1 \otimes \eta)(1 \otimes \Delta)(1 \otimes 1 \otimes \eta)(1 \otimes 1 \otimes f \otimes 1 \otimes 1)(1 \otimes \nabla \otimes 1 \otimes 1) \\
&\quad (1 \otimes \epsilon \otimes 1 \otimes 1)(\nabla \otimes 1)(\epsilon \otimes 1) \\
&= (\eta \otimes 1)(\Delta \otimes 1)(1 \otimes \nabla)(f \otimes 1)((\eta)(\Delta \otimes 1)(1 \otimes \nabla)(1 \otimes \epsilon)) \otimes 1)((1 \otimes \epsilon) \\
&= (1 \otimes \eta)\nabla\Delta(1 \otimes \epsilon)f(\eta \otimes 1)\nabla\Delta(1 \otimes \epsilon) \\
&= id_x id_x f id_y id_y \\
&= f
\end{aligned}$$

$$\begin{aligned}
ff^{(-1)}f &= f(1 \otimes \eta)(1 \otimes \Delta)(1 \otimes f \otimes 1)(\nabla \otimes 1)(\epsilon \otimes 1)f \\
&= (1 \otimes \eta)(1 \otimes \Delta)(f \otimes f \otimes 1)(\nabla \otimes 1)(1 \otimes f)(\epsilon \otimes 1) \\
&= (1 \otimes \eta)(1 \otimes \Delta)(\nabla \otimes 1)(f \otimes f)(\epsilon \otimes 1) \\
&= (1 \otimes \eta)\nabla\Delta(f \otimes f)(\epsilon \otimes 1) \\
&= \Delta(f \otimes f)(\epsilon \otimes 1) \\
&= f\Delta(\epsilon \otimes 1) \\
&= f
\end{aligned}$$

Finally, to show $ff^{(-1)}$ and $gg^{(-1)}$ commute:

$$\begin{aligned}
& f(1 \otimes \eta)(1 \otimes \Delta)(1 \otimes f \otimes 1)(\nabla \otimes 1)(\epsilon \otimes 1)g(1 \otimes \eta)(1 \otimes \Delta)(1 \otimes g \otimes 1)(\nabla \otimes 1)(\epsilon \otimes 1) \\
&= (1 \otimes \eta)(1 \otimes \Delta)(\nabla \otimes 1)(f \otimes 1)(\epsilon \otimes 1)(1 \otimes \eta)(1 \otimes \Delta)(\nabla \otimes 1)(g \otimes 1)(\epsilon \otimes 1) \\
&= (1 \otimes \eta)\nabla\Delta(f \otimes 1)(\epsilon \otimes 1)(1 \otimes \eta)\nabla\Delta(g \otimes 1)(\epsilon \otimes 1) \\
&= \Delta(f \otimes 1)(\epsilon \otimes 1)\Delta(g \otimes 1)(\epsilon \otimes 1) \\
&= \Delta(1 \otimes \Delta)(f \otimes g \otimes 1)(\epsilon \otimes \epsilon \otimes 1) \\
&= \Delta(1 \otimes \Delta)(g \otimes f \otimes 1)(\epsilon \otimes \epsilon \otimes 1) \quad \text{co-commutativity} \\
&= gg^{(-1)}ff^{(-1)}
\end{aligned}$$

□

Chapter 5

Frobenius Algebras and Quantum Computation

Chapter 6

$D[\omega]$ based \dagger categories

6.1 Introduction to synthesis

An important problem in quantum information theory is the decomposition of arbitrary unitary operators into gates from some fixed universal set [18]. Depending on the operator to be decomposed, this may either be done exactly or to within some given accuracy ϵ ; the former problem is known as *exact synthesis* and the latter as *approximate synthesis* [13].

6.2 Algebraic background

We first introduce some notation and terminology, primarily following [13]. Recall that \mathbb{N} is the set of natural numbers including 0, and \mathbb{Z} is the ring of integers. We write $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$ for the ring of integers modulo 2. Let \mathbb{D} be the ring of *dyadic fractions*, defined as $\mathbb{D} = \mathbb{Z}[\frac{1}{2}] = \{\frac{a}{2^n} \mid a \in \mathbb{Z}, n \in \mathbb{N}\}$.

Let $\omega = e^{i\pi/4} = (1+i)/\sqrt{2}$. Note that ω is an 8th root of unity satisfying $\omega^2 = i$ and $\omega^4 = -1$. We will consider three different rings related to ω :

Definition 6.2.1. Consider the following rings. Note that the first two are subrings of the complex numbers, and the third one is not:

- $\mathbb{D}[\omega] = \{a\omega^3 + b\omega^2 + c\omega + d \mid a, b, c, d \in \mathbb{D}\}$.
- $\mathbb{Z}[\omega] = \{a\omega^3 + b\omega^2 + c\omega + d \mid a, b, c, d \in \mathbb{Z}\}$.
- $\mathbb{Z}_2[\omega] = \{p\omega^3 + q\omega^2 + r\omega + s \mid p, q, r, s \in \mathbb{Z}_2\}$.

Note that the ring $\mathbb{Z}_2[\omega]$ only has 16 elements. The laws of addition and multiplication are uniquely determined by the ring axioms and the property $\omega^4 = 1 \pmod{2}$. We call the

elements of $\mathbb{Z}_2[\omega]$ *residues* (more precisely, residue classes of $\mathbb{Z}[\omega]$ modulo 2).

Remark 6.2.2. The ring $\mathbb{D}[\omega]$ is the same as the ring $\mathbb{Z}[\frac{1}{\sqrt{2}}, i]$. However, as already pointed out in [13], the formulation in terms of ω is far more convenient algebraically.

Remark 6.2.3. The ring $\mathbb{Z}[\omega]$ is also called the *ring of algebraic integers* of $\mathbb{D}[\omega]$. It has an intrinsic definition, i.e., one that is independent of the particular presentation of $\mathbb{D}[\omega]$. Namely, a complex number is called an *algebraic integer* if it is the root of some polynomial with integer coefficients and leading coefficient 1. It follows that ω , i , and $\sqrt{2}$ are algebraic integers, whereas, for example, $1/\sqrt{2}$ is not. The ring $\mathbb{Z}[\omega]$ then consists of precisely those elements of $\mathbb{D}[\omega]$ that are algebraic integers.

6.2.1 Conjugate and norm

Remark 6.2.4 (Complex conjugate and norm). Since $\mathbb{D}[\omega]$ and $\mathbb{Z}[\omega]$ are subrings of the complex numbers, they inherit the usual notion of complex conjugation. We note that $\omega^\dagger = -\omega^3$. This yields the following formula:

$$(a\omega^3 + b\omega^2 + c\omega + d)^\dagger = -c\omega^3 - b\omega^2 - a\omega + d. \quad (6.1)$$

Similarly, the sets $\mathbb{D}[\omega]$ and $\mathbb{Z}[\omega]$ inherit the usual norm from the complex numbers. It is given by the following explicit formula, for $t = a\omega^3 + b\omega^2 + c\omega + d$:

$$\|t\|^2 = t^\dagger t = (a^2 + b^2 + c^2 + d^2) + (cd + bc + ab - da)\sqrt{2}. \quad (6.2)$$

Definition 6.2.5 (Weight). For $t \in \mathbb{D}[\omega]$ or $t \in \mathbb{Z}[\omega]$, the *weight* of t is denoted $\|t\|_{\text{weight}}$, and is given by:

$$\|t\|_{\text{weight}}^2 = a^2 + b^2 + c^2 + d^2. \quad (6.3)$$

Note that the square of the norm is valued in $\mathbb{D}[\sqrt{2}]$, whereas the square of the weight is valued in \mathbb{D} . We also extend the definition of norm and weight to vectors in the obvious

way: For $u = (u_j)_j$, we define

$$\|u\|^2 = \sum_j \|u_j\|^2 \quad \text{and} \quad \|u\|_{\text{weight}}^2 = \sum_j \|u_j\|_{\text{weight}}^2.$$

Lemma 6.2.6. *Consider a vector $u \in \mathbb{D}[\omega]^n$. If $\|u\|^2$ is an integer, then $\|u\|_{\text{weight}}^2 = \|u\|^2$.*

Proof. Any $t \in \mathbb{D}[\sqrt{2}]$ can be uniquely written as $t = a + b\sqrt{2}$, where $a, b \in \mathbb{D}$. We can call a the *dyadic part* of t . Now the claim is obvious, because $\|u\|_{\text{weight}}^2$ is exactly the dyadic part of $\|u\|^2$. \square

6.2.2 Denominator exponents

Definition 6.2.7. Let $t \in \mathbb{D}[\omega]$. A natural number $k \in \mathbb{N}$ is called a *denominator exponent* for t if $\sqrt{2}^k t \in \mathbb{Z}[\omega]$. It is obvious that such k always exists. The least such k is called the *least denominator exponent* of t .

More generally, we say that k is a denominator exponent for a vector or matrix if it is a denominator exponent for all of its entries. The least denominator exponent for a vector or matrix is therefore the least k that is a denominator exponent for all of its entries.

Remark 6.2.8. Our notion of least denominator exponent is almost the same as the “smallest denominator exponent” of [13], except that we do not permit $k < 0$.

6.2.3 Residues

Remark 6.2.9. The ring $\mathbb{Z}_2[\omega]$ is not a subring of the complex numbers; rather, it is a quotient of the ring $\mathbb{Z}[\omega]$. Indeed, consider the *parity function* $\overline{(\cdot)} : \mathbb{Z} \rightarrow \mathbb{Z}_2$, which is the unique ring homomorphism. It satisfies $\bar{a} = 0$ if a is even and $\bar{a} = 1$ if a is odd. The parity map induces a surjective ring homomorphism $\rho : \mathbb{Z}[\omega] \rightarrow \mathbb{Z}_2[\omega]$, defined by

$$\rho(a\omega^3 + b\omega^2 + c\omega + d) = \bar{a}\omega^3 + \bar{b}\omega^2 + \bar{c}\omega + \bar{d}.$$

We call ρ the *residue map*, and we call $\rho(t)$ the *residue* of t .

$\rho(t)$	$\rho(\sqrt{2}t)$	$\rho(t^\dagger t)$	$\rho(t)$	$\rho(\sqrt{2}t)$	$\rho(t^\dagger t)$
0000	0000	0000	1000	0101	0001
0001	1010	0001	1001	1111	1010
0010	0101	0001	1010	0000	0000
0011	1111	1010	1011	1010	0001
0100	1010	0001	1100	1111	1010
0101	0000	0000	1101	0101	0001
0110	1111	1010	1110	1010	0001
0111	0101	0001	1111	0000	0000

Table 6.1: Some operations on residues

Convention 6.2.10. Since residues will be important for the constructions of this thesis, we introduce a shortcut notation, writing each residue $p\omega^3 + q\omega^2 + r\omega + s$ as a string of binary digits $pqrs$.

What makes residues useful for our purposes is that many important operations on $\mathbb{Z}[\omega]$ are well-defined on residues. Here, we say that an operation $f : \mathbb{Z}[\omega] \rightarrow \mathbb{Z}[\omega]$ is *well-defined on residues* if for all t, s , $\rho(t) = \rho(s)$ implies $\rho(f(t)) = \rho(f(s))$.

For example, two operations that are obviously well-defined on residues are complex conjugation, which takes the form $(pqrs)^\dagger = rqps$ by ([equation \(6.1\) on page 153](#)), and multiplication by ω , which is just a cyclic shift $\omega(pqrs) = qrsp$. Table [table 6.1](#) shows two other important operations on residues, namely multiplication by $\sqrt{2}$ and the squared norm.

Definition 6.2.11 (k -Residue). Let $t \in \mathbb{D}[\omega]$ and let k be a (not necessarily least) denominator exponent for t . The k -residue of t , in symbols $\rho_k(t)$, is defined to be

$$\rho_k(t) = \rho(\sqrt{2}^k t).$$

Definition 6.2.12 (Reducibility). We say that a residue $x \in \mathbb{Z}_2[\omega]$ is *reducible* if it is of the form $\sqrt{2}y$, for some $y \in \mathbb{Z}_2[\omega]$. Moreover, we say that $x \in \mathbb{Z}_2[\omega]$ is *twice reducible* if it is of the form $2y$, for some $y \in \mathbb{Z}_2[\omega]$.

Lemma 6.2.13. *For a residue x , the following are equivalent:*

- (a) x is reducible;
- (b) $x \in \{0000, 0101, 1010, 1111\}$;
- (c) $\sqrt{2}x = 0000$;
- (d) $x^\dagger x = 0000$.

Moreover, x is twice reducible iff $x = 0000$.

Proof. By inspection of Table [table 6.1 on the preceding page](#). □

Lemma 6.2.14. *Let $t \in \mathbb{Z}[\omega]$. Then $t/2 \in \mathbb{Z}[\omega]$ if and only if $\rho(t)$ is twice reducible, and $t/\sqrt{2} \in \mathbb{Z}[\omega]$ if and only if $\rho(t)$ is reducible.*

Proof. The first claim is trivial, as $\rho(t) = 0000$ if and only if all components of t are even. For the second claim, the left-to-right implication is also trivial: assume $t' = t/\sqrt{2} \in \mathbb{Z}[\omega]$. Then $\rho(t) = \rho(\sqrt{2}t')$, which is reducible by definition. Conversely, let $t \in \mathbb{Z}[\omega]$ and assume that $\rho(t)$ is reducible. Then $\rho(t) \in \{0000, 0101, 1010, 1111\}$, and it can be seen from Table [table 6.1 on the previous page](#) that $\rho(\sqrt{2}t) = 0000$. Therefore, $\sqrt{2}t$ is twice reducible by the first claim; hence t is reducible. □

Corollary 6.2.15. *Let $t \in \mathbb{D}[\omega]$ and let $k > 0$ be a denominator exponent for t . Then k is the least denominator exponent for t if and only if $\rho_k(t)$ is irreducible.*

Proof. Since k is a denominator exponent for t , we have $\sqrt{2}^k t \in \mathbb{Z}[\omega]$. Moreover, k is least if and only if $\sqrt{2}^{k-1} t \notin \mathbb{Z}[\omega]$. By Lemma [6.2.14](#), this is the case if and only if $\rho(\sqrt{2}^k t) = \rho_k(t)$ is irreducible. □

Lemma 6.2.16. *For all a in $\mathbb{Z}[\omega]$, $a + a^t$ is divisible by $\sqrt{2}$ in $\mathbb{Z}[\omega]$.*

Proof. It is sufficient to show this on the generators $\{1, \omega, \omega^2, \omega^3\}$.

1. $1 + 1 = 2$ and $\frac{2}{\sqrt{2}} = \sqrt{2} = \omega - \omega^3$.
2. $\omega + \omega^t = \omega - \omega^3 = \sqrt{2}$.

$$3. \omega^2 + (\omega^2)^t = \omega^2 - \omega^2 = 0.$$

$$4. \omega^3 + (\omega^3)^t = \omega^3 - \omega = -\sqrt{2}.$$

□

Definition 6.2.17. The notions of residue, k -residue, reducibility, and twice-reducibility all extend in an obvious componentwise way to vectors and matrices. Thus, the residue $\rho(u)$ of a vector or matrix u is obtained by taking the residue of each of its entries, and similar for k -residues. Also, we say that a vector or matrix is reducible if each of its entries is reducible, and similarly for twice-reducibility.

Example 6.2.18. Consider the matrix

$$U = \frac{1}{\sqrt{2}^3} \begin{pmatrix} -\omega^3 + \omega - 1 & \omega^2 + \omega + 1 & \omega^2 & -\omega \\ \omega^2 + \omega & -\omega^3 + \omega^2 & -\omega^2 - 1 & \omega^3 + \omega \\ \omega^3 + \omega^2 & -\omega^3 - 1 & 2\omega^2 & 0 \\ -1 & \omega & 1 & -\omega^3 + 2\omega \end{pmatrix}.$$

It has least denominator exponent 3. Its 3-, 4-, and 5-residues are:

$$\begin{aligned} \rho_3(U) &= \begin{pmatrix} 1011 & 0111 & 0100 & 0010 \\ 0110 & 1100 & 0101 & 1010 \\ 1100 & 1001 & 0000 & 0000 \\ 0001 & 0010 & 0001 & 1000 \end{pmatrix}, \\ \rho_4(U) &= \begin{pmatrix} 1010 & 0101 & 1010 & 0101 \\ 1111 & 1111 & 0000 & 0000 \\ 1111 & 1111 & 0000 & 0000 \\ 1010 & 0101 & 1010 & 0101 \end{pmatrix}, \quad \rho_5(U) = 0. \end{aligned}$$

6.3 Exact synthesis of single qubit operators

Matsumoto and Amano [16] showed that every single-qubit Clifford+ T operator can be uniquely written in the following form, which we call the *Matsumoto-Amano normal form*:

$$(T \mid \varepsilon) (HT \mid SHT)^* \mathcal{C}. \tag{6.4}$$

Here, we have used the syntax of regular expressions [12] to denote a set of sequences of operators. The symbol ε denotes the empty sequence (more precisely, the singleton set containing just the empty sequence); if \mathcal{L} and \mathcal{K} are two sets of sequences, then $\mathcal{L} \mid \mathcal{K}$ denotes their union; $\mathcal{L}\mathcal{K}$ denotes the set $\{st \mid s \in \mathcal{L}, t \in \mathcal{K}\}$; \mathcal{L}^* denotes the set $\{s_1 \dots s_n \mid n \geq 0; s_1, \dots, s_n \in \mathcal{L}\}$; and \mathcal{C} denotes any Clifford operator. In words, the Matsumoto-Amano representation of an operator consists of a Clifford operator, followed by any number of *syllables* of the form HT or SHT , followed by an optional syllable T . (We follow the usual convention of multiplying operators right-to-left, so when we say one operator “follows” another, we mean that it appears to its left).

The most important properties of the Matsumoto-Amano decomposition are:

- Existence: all single-qubit Clifford+ T operators can be written in Matsumoto-Amano normal form (moreover, there is an efficient algorithm for converting the operator to normal form);
- Uniqueness: no operator can be written in Matsumoto-Amano normal form in more than one way;
- T -optimality: of all the possible exact decompositions of a given operator into the Clifford+ T set of gates, the Matsumoto-Amano normal form contains the smallest possible number of T -gates.

It is perhaps less well-known that the uniqueness proof given by Matsumoto and Amano yields an efficient *algorithm* for T -optimal exact single-qubit synthesis. One may contrast this, for example, with the recent algorithm by Kliuchnikov et al. [13], which is efficient, but only asymptotically T -optimal. The purpose of this note is to give a detailed presentation of the algorithmic content of Matsumoto and Amano’s result. Along the way, we also simplify Matsumoto and Amano’s proofs, and we give an intrinsic characterization of the Clifford+ T subgroup of $SO(3)$.

6.3.1 Existence

In the following, we will often speak of sequences of operators. For our purposes, a sequence is just an n -tuple. We write st for the operation of concatenating two sequences, and we write ε for the empty sequence. We identify a 1-tuple (A) with the operator A itself. If $s = (A_1, \dots, A_n)$ is a sequence of operators, we write $\llbracket s \rrbracket = A_1 \cdots A_n$ for the product of the operators in the sequence; naturally, $\llbracket \varepsilon \rrbracket = I$. Note that the notation is ambiguous; for example, depending on the context, SHT may denote either the sequence (S, H, T) of 3 operators, or their product, which is a single operator. To alleviate the ambiguity, we assume that everything is a sequence by default, and we write $s \equiv t$ if two sequences are equal as tuples, and $s = t$ if they are equal as operators, i.e., if $\llbracket s \rrbracket = \llbracket t \rrbracket$.

Remark 6.3.1. Consider any operator A in Matsumoto-Amano normal form. If λ is any unit scalar, then λA can clearly also be written in Matsumoto-Amano normal form with the same T -count, namely by multiplying λ into the rightmost Clifford operator. Moreover, if A can be *uniquely* written in Matsumoto-Amano normal form, then the same is true for λA . Therefore, nothing is added or lost to the Matsumoto-Amano normal form whether one allows arbitrary global phases, a suitable discrete set of global phases (for example, powers of $e^{i\pi/4}$), or whether one works modulo global phase. Since it is convenient to work modulo global phase, we do so in the remainder; however, this does not restrict the generality of the results.

Definition 6.3.2. Let \mathcal{C} denote the Clifford group on one qubit, modulo global phases. This group has 24 elements. Let \mathcal{S} be the 8-element subgroup spanned by S and X . Let $\mathcal{C}' = \mathcal{C} \setminus \mathcal{S}$. Let $\mathcal{H} = \{I, H, SH\}$ and $\mathcal{H}' = \{H, SH\}$.

Lemma 6.3.3. *The following hold:*

$$\mathcal{C} = \mathcal{H}\mathcal{S}, \quad (6.5)$$

$$\mathcal{C}' = \mathcal{H}'\mathcal{S}, \quad (6.6)$$

$$\mathcal{S}\mathcal{H}' \subseteq \mathcal{H}'\mathcal{S}, \quad (6.7)$$

$$\mathcal{S}T = T\mathcal{S}, \quad (6.8)$$

$$T\mathcal{S}T = \mathcal{S}. \quad (6.9)$$

Proof. Since \mathcal{S} is an 8-element subgroup of \mathcal{C} , it has three left cosets. They are \mathcal{S} , $H\mathcal{S}$, and $SH\mathcal{S}$. Since \mathcal{C} is the disjoint union of these cosets, (equation (6.5)) and (equation (6.6)) immediately follow. For (equation (6.7)), first notice that $\mathcal{S}\mathcal{S} = \mathcal{S}$, and therefore $\mathcal{S}\mathcal{H}' = \mathcal{S}H \cup \mathcal{S}SH = \mathcal{S}H$. Since $\mathcal{S}H$ is a non-trivial right coset of \mathcal{S} , it follows that $\mathcal{S}H \subseteq \mathcal{C} \setminus \mathcal{S} = \mathcal{C}'$. Combining these facts with (equation (6.6)), we have (equation (6.7)). Finally, the equations (equation (6.8)) and (equation (6.9)) are trivial consequences of the equations $ST = TS$, $XT = TXS$, and $TT = S$. \square

Proposition 6.3.4 (Matsumoto and Amano [16]). *Every single-qubit Clifford+ T operator can be written in Matsumoto-Amano normal form.*

Proof. Let M be a single-qubit Clifford+ T operator. Clearly, M can be written as

$$M = C_n T C_{n-1} \cdots C_1 T C_0, \quad (6.10)$$

for some $n \geq 0$, where $C_0, \dots, C_n \in \mathcal{C}$. First note that if $C_i \in \mathcal{S}$ for any $i \in \{1, \dots, n-1\}$, then we can immediately use (equation (6.9)) to replace TC_iT by a single Clifford operator. This yields a shorter expression of the form (equation (6.10)) for M . We may therefore assume without loss of generality that $C_i \notin \mathcal{S}$ for $i = 1, \dots, n-1$. If $n = 0$, then M is a

Clifford operator, and there is nothing to show. Otherwise, we have

$$M \in \mathcal{C} T \mathcal{C}' \cdots \mathcal{C}' T \mathcal{C} \quad \text{by (equation (6.10))} \quad (6.11)$$

$$= \mathcal{H} \mathcal{S} T \mathcal{H}' \mathcal{S} \cdots \mathcal{H}' \mathcal{S} T \mathcal{C} \quad \text{by (equation (6.5)) and (equation (6.6))} \quad (6.12)$$

$$\subseteq \mathcal{H} T \mathcal{H}' \cdots \mathcal{H}' T \mathcal{C} \quad \text{by (equation (6.7)) and (equation (6.8)).} \quad (6.13)$$

Note how, in the last step, the relations (equation (6.7) on the preceding page) and (equation (6.8) on the previous page) were used to move all occurrences of \mathcal{S} to the right, where they were absorbed into the final \mathcal{C} . It is now trivial to see that every element of (equation (6.13)) can be written in Matsumoto-Amano normal form, finishing the proof. \square

Corollary 6.3.5. *There exists a linear-time algorithm for symbolically reducing any sequence of Clifford+ T operators to Matsumoto-Amano normal form. More precisely, this algorithm runs in time at most $O(n)$, where n is the length of the input sequence.*

Proof. The proof of Proposition 6.3.4 on the previous page already contains an algorithm for reducing any sequence of Clifford+ T operators to Matsumoto-Amano normal form. However, in the stated form, it is perhaps not obvious that the algorithm runs in linear time. Indeed, a naive implementation of the first step would require up to n searches of the entire sequence for a term of the form TST , which can take time $O(n^2)$.

One obtains a linear time algorithm from the following observation: if M is already in Matsumoto-Amano normal form, and A is either a Clifford operator or T , then MA can be reduced to Matsumoto-Amano normal form in constant time. This is trivial when A is a Clifford operator, because it will simply be absorbed into the rightmost Clifford operator of M . In the case where $A = T$, a simple case distinction shows that at most the rightmost 5 elements of MA need to be updated. The normal form of a sequence of operators $A_1 A_2 \dots A_n$ can now be computed by starting with $M = I$ and repeatedly right-multiplying by A_1, \dots, A_n , reducing to normal form after each step. \square

6.3.2 T -Optimality

Corollary 6.3.6. *Let M be an operator in the Clifford+ T group, and assume that M can be written with T -count n . Then there exists a Matsumoto-Amano normal form for M with T -count at most n .*

Proof. This is an immediate consequence of the proof of Proposition 6.3.4 on page 160, because the reduction from (equation (6.10) on page 160) to (equation (6.13) on the preceding page) does not increase the T -count. \square

6.3.3 Uniqueness

Theorem 6.3.7 (Matsumoto and Amano [16]). *If M and N are two different Matsumoto-Amano normal forms, then they describe different operators.*

Recall that each single-qubit unitary operator (modulo global phase) can be uniquely represented as a rotation on the Bloch sphere, or equivalently, as an element of $SO(3)$, the real orthogonal 3×3 matrices with determinant 1. The relationship between an operator $U \in U(2)$ and its Bloch sphere representation $\hat{U} \in SO(3)$ is given by

$$\hat{U} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \iff U(xX + yY + zZ)U^\dagger = x'X + y'Y + z'Z, \quad (6.14)$$

where X , Y , and Z are the Pauli operators. The Bloch sphere representations of the operators H , S , and T are:

$$\hat{H} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \hat{S} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \hat{T} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}. \quad (6.15)$$

Remark 6.3.8. The Bloch sphere representation of any scalar is the identity matrix. The Bloch sphere representations of the 24 Clifford operators (modulo phase) are precisely those elements of $SO(3)$ that can be written with matrix entries in $\{-1, 0, 1\}$; these are exactly the 24 symmetries of the cube $\{(x, y, z) \mid -1 \leq x, y, z \leq 1\}$.

Definition 6.3.9. Recall that \mathbb{N} denotes the natural numbers including 0; \mathbb{Z} denotes the integers; and \mathbb{Z}_2 denotes the integers modulo 2. We define three subrings of the real numbers:

- $\mathbb{D} = \mathbb{Z}[\frac{1}{2}] = \{\frac{a}{2^n} \mid a \in \mathbb{Z}, n \in \mathbb{N}\}$. This is the ring of *dyadic fractions*.
- $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\}$. This is the ring of *quadratic integers* with radicand 2.
- $\mathbb{D}[\sqrt{2}] = \mathbb{Z}[\frac{1}{\sqrt{2}}] = \{r + s\sqrt{2} \mid r, s \in \mathbb{D}\}$.

We will also need a fourth ring, which is not a subring of the real numbers.

- $\mathbb{Z}_2[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}_2\}$.

Note that the ring $\mathbb{Z}_2[\sqrt{2}]$ has only 4 elements; they are residue classes modulo 2 of the ring $\mathbb{Z}[\sqrt{2}]$. For brevity, we refer to the elements of $\mathbb{Z}_2[\sqrt{2}]$ as *residues*.

Definition 6.3.10 (Residue¹ and parity). Consider the unique ring homomorphism $\mathbb{Z} \rightarrow \mathbb{Z}_2$, mapping $a \in \mathbb{Z}$ to $\bar{a} \in \mathbb{Z}_2$, where $\bar{a} = 0$ if a is even and $\bar{a} = 1$ if a is odd. This induces a surjective ring homomorphism $\rho : \mathbb{Z}[\sqrt{2}] \rightarrow \mathbb{Z}_2[\sqrt{2}]$, defined by $\rho(a + b\sqrt{2}) = \bar{a} + \bar{b}\sqrt{2}$. For any given $x \in \mathbb{Z}[\sqrt{2}]$, we refer to $\rho(x)$ as the *residue of x* .

Moreover, consider the ring homomorphism $p : \mathbb{Z}[\sqrt{2}] \rightarrow \mathbb{Z}_2$ given by $p(a + b\sqrt{2}) = \bar{a}$. We refer to $p(x)$ as the *parity of x* .

Definition 6.3.11 (Denominator exponent). For every element $q \in \mathbb{D}[\sqrt{2}]$, there exists some natural number $k \geq 0$ such that $\sqrt{2}^k q \in \mathbb{Z}[\sqrt{2}]$, or equivalently, such that q can be written as $\frac{x}{\sqrt{2}^k}$, for some quadratic integer x . Such k is called a *denominator exponent* for q . The least such k is called the *least denominator exponent of q* .

More generally, we say that k is a denominator exponent for a vector or matrix if it is a denominator exponent for all of its entries. The least denominator exponent for a vector or matrix is therefore the least k that is a denominator exponent for all of its entries.

¹I don't think we need residue here, which is good as it will then be used only in section 7, the $U(2)$ case

$$\begin{array}{ccc}
\text{Start: } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \xrightarrow{\mathfrak{C}} & \\
\downarrow \begin{matrix} k++ \\ T \end{matrix} & & \\
\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \xleftarrow[k++]{T} & \\
\begin{matrix} H \downarrow \\ \uparrow T_{k++} \end{matrix} & & \\
\begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} & \xrightarrow{S} & \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}
\end{array}$$

Figure 6.1: The action of Matsumoto-Amano normal forms on k -parities. All matrices are written modulo the right action of the Clifford group, i.e., modulo a permutation of the columns.

Definition 6.3.12 (k -parity). Let k be a denominator exponent for $q \in \mathbb{D}[\sqrt{2}]$. We define the k -residue of q , in symbols $\rho_k(q) \in \mathbb{Z}_2[\sqrt{2}]$, and the k -parity of q , in symbols $p_k(q) \in \mathbb{Z}_2$, by

$$\rho_k(q) = \rho(\sqrt{2}^k q), \quad p_k(q) = p(\sqrt{2}^k q).$$

The k -residue and k -parity of a vector or matrix are defined componentwise.

Remark 6.3.13. Let C be any Clifford operator, and \hat{C} its Bloch sphere representation. As noted above, the matrix entries of \hat{C} are in $\{-1, 0, 1\}$; it follows that \hat{C} has denominator exponent 0. In particular, it follows that multiplication by \hat{C} is a well-defined operation on parity matrices: for any 3×3 -matrix U with entries in $\mathbb{Z}_2[\sqrt{2}]$, we define $U \bullet C := U \cdot p(\hat{C})$. This defines a right action of the Clifford group on the set of parity matrices.

Definition 6.3.14. If G is any subgroup of the Clifford group, we define \sim_G to be the equivalence relation induced by this right action, i.e., for parity matrices U, V , we write $U \sim_G V$ if there exists some $C \in G$ such that $V = U \bullet C$. In case $G = \mathfrak{C}$ is the entire Clifford group, $U \sim_{\mathfrak{C}} V$ holds if and only if U and V differ by a permutation of columns.

Lemma 6.3.15. *Let M be a Matsumoto-Amano normal form, and $\hat{M} \in SO(3)$ the Bloch sphere operator of M . Let k be the least denominator exponent of \hat{M} . Then exactly one of the following holds:*

- $k = 0$, and M is a Clifford operator.
- $k > 0$, $p_k(\hat{M}) \sim_{\mathcal{C}} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, and the leftmost syllable in M is T .
- $k > 0$, $p_k(\hat{M}) \sim_{\mathcal{C}} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$, and the leftmost syllable in M is HT .
- $k > 0$, $p_k(\hat{M}) \sim_{\mathcal{C}} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$, and the leftmost syllable in M is SHT .

Moreover, the T -count of M is equal to k .

Proof. By induction on the length of the Matsumoto-Amano normal form M . Figure [figure 6.1 on the previous page](#) shows the action of Matsumoto-Amano operators on parity matrices. Each vertex represents a $\sim_{\mathcal{C}}$ -equivalence class of k -parities. The vertex labelled “Start” represents the empty Matsumoto-Amano normal form, i.e., the identity operator. Each arrow represents left multiplication by the relevant operator, i.e., a Clifford operator, T , H , or S . Thus, each Matsumoto-Amano normal form, read from right to left, gives rise to a unique path in the graph of Figure [figure 6.1 on the preceding page](#). The label $k++$ on an arrow indicates that the least denominator exponent increases by 1. The claims of the lemma then immediately follow from Figure [figure 6.1 on the previous page](#). \square

Proof of Theorem 6.3.7 on page 162. This is an immediate consequence of Lemma 6.3.15. Indeed, suppose that M and N are two Matsumoto-Amano normal forms describing the same Bloch sphere operator U . We show that $M = N$ by induction on the length of M . Let k be the least denominator exponent of U . If $k = 0$, then by Lemma 6.3.15, both M and N

are Clifford operators; since they describe the same Bloch sphere operator, they differ only by a phase.² If $k > 0$, then by Lemma 6.3.15, the Matsumoto-Amano normal forms M and N have the same leftmost operator (either T , H , or S), and the claim follows by induction hypothesis. \square

6.3.4 The Matsumoto-Amano decomposition algorithm

As an immediate consequence of Lemma 6.3.15 on page 164, we can an efficient algorithm for calculating the Matsumoto-Amano normal form of any Clifford+ T operator, given as a matrix.

Theorem 6.3.16. *Let $U \in SO(3)$ be the Bloch sphere representation of some Clifford+ T operator. Let k be the least denominator exponent of U . Then the Matsumoto-Amano normal form M of U can be efficiently computed with $O(k)$ arithmetic operations.*

Proof. By assumption, U is the Bloch sphere representation of some Clifford+ T operator. Let M be the unique Matsumoto-Amano normal form of this operator³. Note that, by Lemma 6.3.15 on page 164, the T -count of M is k . We compute M recursively. If $k = 0$, then by Lemma 6.3.15 on page 164, M is a Clifford operator; it can be determined from the matrix U in constant time. If $k > 0$, we compute $p_k(U)$, which must be one of the three cases listed in Lemma 6.3.15 on page 164. This determines whether the leftmost syllable of M is T , HT , or SHT . Let N be this syllable, so that $M = NM'$, for some Matsumoto-Amano normal form M' . Then M' can be recursively computed as the Matsumoto-Amano normal form of $U' = \hat{N}^{-1}U$; moreover, since M' has T -count $k - 1$, the recursion terminates after k steps. Since each induction step only requires a constant number of arithmetic operations, the total number of operations is $O(k)$. \square

²Todo: fix the treatment of phases

³Todo: treat phase correctly

6.3.5 A characterization of Clifford+ T on the Bloch sphere

Lemma 6.3.17. *Let $U \in SO(3)$ be an orthogonal matrix with entries in $\mathbb{D}[\sqrt{2}]$. Let k be a denominator exponent of U , and let v_1, v_2, v_3 be the columns of U , with*

$$v_j = \frac{1}{\sqrt{2}^k} \begin{pmatrix} a_j + b_j\sqrt{2} \\ c_j + d_j\sqrt{2} \\ e_j + f_j\sqrt{2} \end{pmatrix},$$

for $a_j, \dots, f_j \in \mathbb{Z}$. Then for all $j, \ell \in \{1, 2, 3\}$,

$$a_j b_\ell + b_j a_\ell + c_j d_\ell + d_j c_\ell + e_j f_\ell + f_j e_\ell = 0 \quad (6.16)$$

and

$$a_j a_\ell + c_j c_\ell + e_j e_\ell + 2(b_j b_\ell + d_j d_\ell + f_j f_\ell) = 2^k \langle v_j, v_\ell \rangle. \quad (6.17)$$

In particular, we have, for all $j \in \{1, 2, 3\}$,

$$a_j b_j + c_j d_j + e_j f_j = 0 \quad (6.18)$$

and

$$a_j^2 + c_j^2 + e_j^2 + 2(b_j^2 + d_j^2 + f_j^2) = 2^k. \quad (6.19)$$

Proof. Computing the inner product, we have

$$\begin{aligned} \langle v_j, v_\ell \rangle = & \frac{1}{2^k} (a_j a_\ell + c_j c_\ell + e_j e_\ell + 2(b_j b_\ell + d_j d_\ell + f_j f_\ell) \\ & + \sqrt{2}(a_j b_\ell + b_j a_\ell + c_j d_\ell + d_j c_\ell + e_j f_\ell + f_j e_\ell)). \end{aligned} \quad (6.20)$$

Since $U^\dagger U = I$, we have $\langle v_j, v_j \rangle = 1$, and $\langle v_j, v_\ell \rangle = 0$ when $\ell \neq j$. Therefore, the coefficient of $\sqrt{2}$ in equation (6.20) must be zero, proving (6.16) and (6.17).

Equations (6.18) and (6.19) immediately follow by letting $j = \ell$. \square

Remark 6.3.18. In Lemma 6.3.17 on the preceding page, we have worked with columns v_j of the matrix U . But since U is orthogonal, the analogous properties also hold for the rows of U .

Lemma 6.3.19. Let $U \in SO(3)$ be an orthogonal matrix with entries in $\mathbb{D}[\sqrt{2}]$, and with least denominator exponent $k = 0$. Then U the Bloch sphere representation of some Clifford operator.

Proof. Let v_j be any column of U , with the notation of Lemma 6.3.17 on the previous page. By equation (6.19) on the preceding page, we have $a_j^2 + c_j^2 + e_j^2 + 2(b_j^2 + d_j^2 + f_j^2) = 1$. Since each summand is a positive integer, we must have $b_j, d_j, f_j = 0$, and exactly one of a_j, c_j or $e_j = \pm 1$, for each $j = 1, 2, 3$. Therefore, all the matrix entries are in $\{-1, 0, 1\}$, and the claim follows by Remark 6.3.8 on page 162. \square

Lemma 6.3.20. Let $U \in SO(3)$ be an orthogonal matrix with entries in $\mathbb{D}[\sqrt{2}]$, and let k be the least denominator exponent of U . If $k = 0$, then $p_k(U) \sim_{\mathcal{C}} M_1$. If $k > 0$, then $p_k(U) \sim_{\mathcal{C}} M$ for some $M \in \{M_T, M_H, M_S\}$, where

$$M_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad M_T = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad M_H = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \quad M_S = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}.$$

Proof. First consider the case $k = 0$. Let v_j be any column of U , with the notation of Lemma 6.3.17 on the preceding page. By equation (6.19) on the previous page, we have $a_j^2 + c_j^2 + e_j^2 + 2(b_j^2 + d_j^2 + f_j^2) = 1$. Since each summand is a positive integer, we must have $b_j, d_j, f_j = 0$, and exactly one of a_j, c_j or $e_j = \pm 1$, for each $j = 1, 2, 3$. Noting that the columns of U are orthogonal, we see that $p_k(U) \sim_{\mathcal{C}} M_1$.

Now consider the case $k > 0$. Let v_j be any row or column of U , with the notation of Lemma 6.3.17 on the preceding page. By equation (6.19) on the previous page, it follows that $a_j^2 + c_j^2 + e_j^2$ is even, and therefore an even number of a_j, c_j , and e_j have parity 1. Therefore, each row or column of $p_k(U)$ has an even number of 1's. Moreover, since k is the least

denominator exponent of U , $p_k(U)$ has at least one non-zero entry. Modulo a permutation of columns, this leaves exactly four possibilities for $p_k(U)$:

$$(a) \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (b) \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \quad (c) \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \quad (d) \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

In cases (a)–(c), we are done. Case (d) is impossible because it implies that $a_1a_2 + c_1c_2 + e_1e_2$ is odd, contradicting the fact that it is even by [equation \(6.17\) on page 167](#). \square

Lemma 6.3.21. *Let $U \in SO(3)$ be an orthogonal matrix with entries in $\mathbb{D}[\sqrt{2}]$, and with least denominator exponent $k > 0$. Then there exists $N \in \{T, HT, SHT\}$ such that the least denominator exponent of $\hat{N}^{-1}U$ is $k - 1$.*

Proof. By [Lemma 6.3.20 on the previous page](#), we know that $p_k(U) \sim_{\mathbb{C}} M$, for some $M \in \{M_T, M_H, M_S\}$. We consider each of these cases.

1. $p_k(U) \sim_{\mathbb{C}} M_T$. By assumption, U has two columns v with $p_k(v) = (1, 1, 0)^T$.

Let

$$v = \frac{1}{\sqrt{2}^k} \begin{pmatrix} a + b\sqrt{2} \\ c + d\sqrt{2} \\ e + f\sqrt{2} \end{pmatrix}$$

be any such column. By [equation \(6.18\) on page 167](#), we have $ab + cd + ef = 0$.

Since $\bar{e} = 0$, we have $\bar{a}\bar{b} + \bar{c}\bar{d} = 0$. Since $\bar{a} = \bar{c} = 1$, we can conclude $\bar{b} + \bar{d} = 0$.

Applying \hat{T}^{-1} to v , we compute:

$$\begin{aligned} \hat{T}^{-1}v &= \frac{1}{\sqrt{2}^{k+1}} \begin{bmatrix} c + a & + & (d + b)\sqrt{2} \\ c - a & + & (d - b)\sqrt{2} \\ e\sqrt{2} & + & 2f \end{bmatrix} = \frac{1}{\sqrt{2}^{k-1}} \begin{bmatrix} \frac{c+a}{2} & + & \frac{d+b}{\sqrt{2}} \\ \frac{c-a}{2} & + & \frac{d-b}{\sqrt{2}} \\ \frac{e}{\sqrt{2}} & + & f \end{bmatrix} \\ &= \frac{1}{\sqrt{2}^{k-1}} \begin{bmatrix} a' & + & b'\sqrt{2} \\ c' & + & d'\sqrt{2} \\ f & + & e'\sqrt{2} \end{bmatrix} \end{aligned}$$

where $a' = \frac{c+a}{2}, b' = \frac{d+b}{2}, c' = \frac{c-a}{2}, d' = \frac{d-b}{2}$ and $e' = \frac{e}{2}$ are all integers. Hence, $k-1$ is a denominator exponent of $\hat{T}^{-1}v$. Moreover, since $a' + c' = c$ is odd, one of a' and c' is odd, proving that $k-1$ is the least denominator exponent of $\hat{T}^{-1}v$.

Now consider the third column w of U , where $p_k(w) = (0, 0, 0)^T$. Then $k-1$ is a denominator exponent of w , so that k is a denominator exponent for $\hat{T}^{-1}w$.

Let

$$p_k(\hat{T}^{-1}w) = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

As the least denominator exponent of the other two column of $p_k(\hat{T}^{-1}U)$ is $k-1$, we have

$$p_k(\hat{T}^{-1}U) \sim_{\mathbb{C}} \begin{bmatrix} 0 & 0 & x \\ 0 & 0 & y \\ 0 & 0 & z \end{bmatrix}.$$

But $\hat{T}^{-1}U$ is orthogonal, so by [equation \(6.19\) on page 167](#), applied to each row of $\hat{T}^{-1}U$, we conclude that $x = y = z = 0$. It follows that the least denominator exponent of $\hat{T}^{-1}U$ is $k-1$.

2. $p_k(U) \sim_{\mathbb{C}} M_H$. In this case, $p_k(\hat{H}^{-1}U) \sim_{\mathbb{C}} p(\hat{H}^{-1}M_H) = M_T$. We then continue as in [case 1 on the preceding page](#).
3. $p_k(U) \sim_{\mathbb{C}} M_S$. In this case, $p_k(\hat{H}^{-1}\hat{S}^{-1}U) \sim_{\mathbb{C}} p(\hat{H}^{-1}\hat{S}^{-1}M_S) = M_T$. We then continue as in [case 1 on the previous page](#). □

Combining [Lemmas 6.3.19 on page 168](#) and [6.3.20 on page 168](#), we easily get the following result:

Theorem 6.3.22. *Let $U \in SO(3)$ be an orthogonal matrix. Then U is the Bloch sphere representation of some Clifford+T operator M if and only if the entries of U are in the ring $\mathbb{D}[\sqrt{2}]$.*

Proof. The “only if” direction is trivial, since all the generators of the Clifford+ T group have this property (see [equation \(6.15\)](#)). To prove the “if” direction, let k be the least denominator exponent of U . We proceed by induction on k . If $k = 0$, by [Lemma 6.3.19](#), U is the Bloch sphere representation of some Clifford operator, and therefore a Clifford+ T operator. If $k > 0$, then by [Lemma 6.3.20](#), we can write $U = \hat{N}U'$, where $N \in \{T, HT, SHT\}$ and U' has least denominator exponent $k - 1$. By induction hypothesis, U' is a Clifford+ T operator, and therefore so is U . \square

Remark 6.3.23. Combining this result with the algorithm of [Theorem 6.3.16 on page 166](#), we have a linear-time algorithm for computing the Matsumoto-Amano normal form of any unitary operator $U \in SO(3)$ with entries in $\mathbb{D}[\sqrt{2}]$.

Corollary 6.3.24 (Kliuchnikov et al. [\[13\]](#)). *Let $U \in U(2)$ be a unitary matrix. Then U is a Clifford+ T operator if and only if the matrix entries of U are in the ring $\mathbb{D}[\sqrt{2}, i]$.*

Proof. Again, the “only if” direction is trivial, as it is true for the generators. For the “if” direction, it suffices to note that, by [equation \(6.14\) on page 162](#), whenever U takes its entries in $\mathbb{D}[\sqrt{2}, i]$, then \hat{U} takes its entries in $\mathbb{D}[\sqrt{2}]$.⁴ \square

[Corollary 6.3.24](#) was first proved by Kliuchnikov et al. [\[13\]](#), using a direct method (i.e., not going via the Bloch sphere representation). It is interesting to note that [Theorem 6.3.22 on the previous page](#) is stronger than [Corollary 6.3.24](#), in the sense that the Theorem obviously implies the Corollary, whereas it is not a priori obvious that the Corollary implies the Theorem.

6.3.6 Alternative normal forms

With the exception of the left-most and right-most gates, the Matsumoto-Amano normal form uses syllables of the form HT and SHT . It is of course possible to use different sets of syllables instead.

⁴Todo: treat phase correctly; also introduce the ring $\mathbb{D}[\sqrt{2}, i]$ at some appropriate time

E - T normal form

Consider the Clifford operator

$$E = HS^3\omega^3 = \frac{1}{2} \begin{pmatrix} -1+i & 1+i \\ -1+i & -1-i \end{pmatrix}.$$

It has the following properties:

$$E^3 = I, \quad EXE^{-1} = Y, \quad EYE^{-1} = Z, \quad EZE^{-1} = X.$$

The operator E serves as a convenient operator for switching between the X -, Y -, and Z -bases. On the Bloch sphere, it represents a rotation by 120 degrees about the axis $(1, 1, 1)^T$:

$$\hat{E} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

The operators E and E^2 have properties analogous to H and SH . Specifically, if we let $\mathcal{H} = \{I, E, E^2\}$ and $\mathcal{H}' = \{E, E^2\}$, then the properties of Lemma 6.3.3 are satisfied. The proofs of Proposition 6.3.4 and Corollary 6.3.5 only depend on these properties, and the uniqueness proof (Theorem 6.3.7 on page 162) also goes through without significant changes. We therefore have:

Proposition 6.3.25 (E - T normal form). *Every single-qubit Clifford+ T operator can be uniquely written in the form*

$$(T \mid \varepsilon) (ET \mid E^2T)^* \mathcal{C}. \quad (6.21)$$

Moreover, this normal form has minimal T -count, and there exists a linear-time algorithm for symbolically reducing any sequence of Clifford+ T operators to this normal form.

T_x - T_y - T_z normal form

It is plain to see that every syllable of the E - T normal form (except perhaps the first or last one) consists of a 45 degree z -rotation, followed by a basis change that rotates either the x - or y -axis into the z -position. Abstracting away from these basis changes, the entire

normal form can therefore be regarded as a sequence of 45-degree rotations about the x -, y -, and z -axes. More precisely, let us define variants of the T -gate that rotate about the three different axes:

$$T_x = ETE^2,$$

$$T_y = E^2TE,$$

$$T_z = T.$$

Using the commutativities $ET_x = T_yE$, $ET_y = T_zE$, and $ET_z = T_xE$, it is then clear that every expression of the form ([equation \(6.21\) on the preceding page](#)) can be uniquely rewritten as a sequence of T_x , T_y , and T_z rotations, with no repeated symbol, followed by a Clifford operator. This can be easily proved by induction, but is best seen in an example:

$$\begin{aligned} TETETE^2TEC &= T_zET_zET_zE^2T_zEC \\ &\rightarrow T_zT_xE^2T_zE^2T_zEC \\ &\rightarrow T_zT_xT_yE^4T_zEC \\ &\rightarrow T_zT_xT_yET_zEC \\ &\rightarrow T_zT_xT_yT_xE^2C \\ &\rightarrow T_zT_xT_yT_xC'. \end{aligned}$$

We have:

Proposition 6.3.26 (T_x - T_y - T_z normal form). *Every single-qubit Clifford+ T operator can be uniquely written in the form*

$$T_{r_1}T_{r_2}\dots T_{r_n}C,$$

where $n \geq 0$, $r_1, \dots, r_n \in \{x, y, z\}$, and $r_i \neq r_{i+1}$ for all $i \leq n-1$. We define the T -count of such an expression to be n ; then this normal form has minimal T -count. Moreover, there exists a linear-time algorithm for symbolically reducing any sequence of Clifford+ T operators to this normal form.

The T_x - T_y - T_z normal form is, in a sense, the most “canonical” one of the normal forms considered here; it also explains why T -count is an appropriate measure of the size of a

Clifford+ T operator. In a physical quantum computer with error correction, there is in general no reason to expect the T_z gate to be more privileged than the T_x or T_y gates; one may imagine that it would be efficient for a quantum computer to provide all three T -gates as primitive logical operations.

Bocharov-Svore normal form

Bocharov and Svore [8, Proposition 1] consider the following normal form for single-qubit Clifford+ T circuits:

$$(H \mid \varepsilon)(TH \mid SHTH)^* \mathcal{C}. \quad (6.22)$$

This normal form is not unique; for example, $H.H$ and I are two different normal forms denoting the same operator, as are $SHTH.Z$ and $H.SHTH$. (Here we have used a dot to delimit syllables; this is for readability only). Recall that two regular expressions are *equivalent* if they define the same set of strings. Using laws of regular expressions, we can equivalently rewrite (equation (6.22)) as

$$((\epsilon \mid T \mid SHT)(HT \mid HSHT)^* HC) \mid \mathcal{C}. \quad (6.23)$$

Since HC is just a redundant way to write a Clifford operator, we can simplify it to \mathcal{C} ; moreover, in this case, $\epsilon\mathcal{C}$ and \mathcal{C} are the same, so (equation (6.23)) simplifies to

$$(\epsilon \mid T \mid SHT)(HT \mid HSHT)^* \mathcal{C}. \quad (6.24)$$

Moreover, since $SHT = HSHT.X$, any expression starting with SHT can be rewritten as one starting with $HSHT$, so the SHT syllable is redundant and we can eliminate it:

$$(\epsilon \mid T)(HT \mid HSHT)^* \mathcal{C}. \quad (6.25)$$

Let us say that an operator is in *Bocharov-Svore normal form* if it is written in the form (equation (6.25)). This version of the Bocharov-Svore normal form is indeed unique; note that it is almost the same as the Matsumoto-Amano normal form, except that the syllable

SHT has been replaced by $HSHT$. Since the set $\mathcal{H} = \{I, H, HSH\}$ satisfies Lemma 6.3.3 on page 159, existence, uniqueness, T -optimality, and efficiency are proved in the same way as for the Matsumoto-Amano and E - T normal forms.

Bocharov and Svore [8, Prop.2] also consider a second normal form, which has Clifford operators on both sides, but the first four interior syllables restricted to TH :

$$\mathcal{C}(\epsilon \mid TH \mid (TH)^2 \mid (TH)^3 \mid (TH)^4(TH \mid SHTH)^*)\mathcal{C} \quad (6.26)$$

However, this normal form is not at all unique; for instance, $Z.TH$ and $TH.X$ denote the same operator, as do $YS.TH.TH$ and $TH.TH.X\omega$.

6.3.7 Matsumoto-Amano normal forms and $U(2)$

Example 6.3.27. Consider the matrix

$$U = \frac{1}{\sqrt{2}^3} \begin{pmatrix} \omega^2 + \omega & -2\omega^3 + \omega^2 + \omega \\ \omega^3 - 2\omega^2 - 1 & -\omega^3 + 1 \end{pmatrix}.$$

It has least denominator exponent 3. Its 3-, 4-, and 5-residues are:

$$\rho_3(U) = \begin{pmatrix} 0110 & 0110 \\ 1001 & 1001 \end{pmatrix}, \quad \rho_4(U) = \begin{pmatrix} 1111 & 1111 \\ 1111 & 1111 \end{pmatrix}, \quad \rho_5(U) = 0.$$

Definition 6.3.28. Let \mathcal{S}^ω be the 64-element subgroup of the Clifford group in $U(2)$ spanned by S, X and ω . Then $\sim_{\mathcal{S}^\omega}$ defined by right multiplication by \mathcal{S}^ω is an equivalence relation on the residue matrices of operators in the Clifford+ T group.

Remark 6.3.29. The equivalence relation $\sim_{\mathcal{S}^\omega}$ is characterized by the following operations:

1. “Rotating” all of the entries in the matrix by 1,2 or 3 positions. This corresponds to multiplication by a power of ω .
2. Swapping the two columns. This corresponds to the right action of X .
3. “Rotating” the entries of the second column by two positions. This corresponds to the right action of S .

Remark 6.3.30. The residue matrices in figure [figure 6.2 on the previous page](#) are modulo $\sim_{S\omega}$.

Lemma 6.3.31. *Let $k \geq 2$ and $b, d \in \mathbb{Z}[\omega]$. Then $\frac{1}{\sqrt{2}^k} \begin{pmatrix} 1 + 2b \\ 1 + 2d \end{pmatrix}$ is not a unit vector.*

Proof. Suppose otherwise, then

$$\begin{aligned} 2^k &= 1 + 2(b + b^t) + 4bb^t + \\ &\quad 1 + 2(d + d^t) + 4dd^t, \end{aligned}$$

so $2 = 2^k - 2(b + b^t) - 2(d + d^t) - 4(bb^t + dd^t)$. By lemma [6.2.16 on page 156](#), the right hand side of this is divisible in $\mathbb{Z}[\omega]$ by $2\sqrt{2}$, while the left hand side is not. Thus we have a contradiction. \square

Lemma 6.3.32. *Given a unitary matrix $U \in \mathbb{D}[\omega]^{2 \times 2}$ with least denominator exponent $k \geq 2$, such that:*

$$\begin{aligned} 1. \quad \rho_{k+1}(U) &= \begin{bmatrix} 0101 & 0101 \\ 0101 & 0101 \end{bmatrix} \text{ and} \\ 2. \quad \rho_k(HU) &= \begin{bmatrix} 0011 & 0011 \\ 0110 & 0110 \end{bmatrix}, \end{aligned}$$

$$\text{then, } \rho_k(U) = \begin{bmatrix} 0010 & 1101 \\ 1101 & 0010 \end{bmatrix}$$

Proof. Referencing Table [table 6.1 on page 155](#), we see the first condition limits the possible choices for the entries of $\rho_k(U)$ to the set $\{0010, 0111, 1000, 1101\}$. The second condition implies that $\rho_{k+1}(HU)$ is reducible and in fact that each entry is 1111. This means each column of U must be either $[0010, 1101]^t$, $[1101, 0010]^t$, $[0111, 1000]^t$ or $[1000, 0111]^t$. As we are considering equivalence classes, we can assume without loss of generality, that the columns

are in $\{[0010, 1101]^t, [1101, 0010]^t\}$. But by Lemma 6.3.31 on the preceding page, we can not have a row like $[0010, 0010]$, therefore $U = \begin{bmatrix} 0010 & 1101 \\ 1101 & 0010 \end{bmatrix}$. \square

Corollary 6.3.33. *In Figure figure 6.2 on page 176, the transitions labelled with “Reduce” are correct.*

Proof. For the left most reduce, we directly apply Lemma 6.3.32 on the preceding page and then note that

$$\begin{bmatrix} 0001 & 1110 \\ 1110 & 0001 \end{bmatrix} \sim_{S^\omega} \begin{bmatrix} 0010 & 1101 \\ 1101 & 0010 \end{bmatrix}.$$

For the rightmost reduce, the same argument as shown in Lemma 6.3.32 on the previous page can be applied to the preconditions of this reduce, giving us a similar result. \square

6.4 Exact synthesis of multi-qubit operators

Here, we focus on the problem of exact synthesis for n -qubit operators, using the Clifford+ T universal gate set. Recall that the Clifford group on n qubits is generated by the Hadamard gate H , the phase gate S , the controlled-not gate, and the scalar $\omega = e^{i\pi/4}$ (one may allow arbitrary unit scalars, but it is not convenient for our purposes to do so). It is well-known that one obtains a universal gate set by adding the non-Clifford operator T [18].

$$\begin{aligned} \omega = e^{i\pi/4}, \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \\ CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}. \end{aligned} \tag{6.27}$$

In addition to the Clifford+ T group on n qubits, as defined above, we also consider the slightly larger group of Clifford+ T operators “with ancillas”. We say that an n -qubit

operator U is a Clifford+ T operator *with ancillas* if there exists $m \geq 0$ and a Clifford+ T operator U' on $n + m$ qubits, such that $U'(|\phi\rangle \otimes |0\rangle) = (U|\phi\rangle) \otimes |0\rangle$ for all n -qubit states $|\phi\rangle$.

Kliuchnikov, Maslov, and Mosca [13] showed that a single-qubit operator U is in the Clifford+ T group if and only if all of its matrix entries belong to the ring $\mathbb{Z}[\frac{1}{\sqrt{2}}, i]$. They also showed that the Clifford+ T groups “with ancillas” and “without ancillas” coincide for $n = 1$, but not for $n \geq 2$. Moreover, Kliuchnikov et al. conjectured that for all n , an n -qubit operator U is in the Clifford+ T group with ancillas if and only if its matrix entries belong to $\mathbb{Z}[\frac{1}{\sqrt{2}}, i]$. They also conjectured that a single ancilla qubit is always sufficient in the representation of a Clifford+ T operator with ancillas. This section of the thesis will prove these conjectures. In particular, this yields an algorithm for exact Clifford+ T synthesis of n -qubit operators. We also obtain a characterization of the Clifford+ T group on n qubits without ancillas.

It is important to note that, unlike in the single-qubit case, the circuit synthesized here are not in any sense canonical, and very far from optimal. Thus, the question of *efficient* synthesis is not addressed here.

6.4.1 Decomposition into two-level matrices

Recall that a *two-level matrix* is an $n \times n$ -matrix that acts non-trivially on at most two vector components [18]. If

$$U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is a 2×2 -matrix and $j \neq \ell$, we write $U_{[j,\ell]}$ for the two-level $n \times n$ -matrix defined by

$$U_{[j,\ell]} = \begin{matrix} & \dots & j & \dots & \ell & \dots \\ \vdots & \left(\begin{array}{c|c|c|c|c} I & & & & \\ \hline & a & & b & \\ \hline & & I & & \\ \hline & c & & d & \\ \hline & & & & I \end{array} \right) & \\ j & & & & & \\ \vdots & & & & & \\ \ell & & & & & \\ \vdots & & & & & \end{matrix},$$

and we say that $U_{[j,\ell]}$ is a two-level matrix *of type U*. Similarly, if a is a scalar, we write $a_{[j]}$ for the one-level matrix

$$a_{[j]} = \begin{matrix} & \dots & j & \dots \\ \vdots & \left(\begin{array}{c|c|c} I & & \\ \hline & a & \\ \hline & & I \end{array} \right) & \\ j & & & \\ \vdots & & & \end{matrix},$$

and we say that $a_{[j]}$ is a one-level matrix *of type a*.

Lemma 6.4.1 (Row operation). *Let $u = (u_1, u_2)^T \in \mathbb{D}[\omega]^2$ be a vector with denominator exponent $k > 0$ and k -residue $\rho_k(u) = (x_1, x_2)$, such that $x_1^\dagger x_1 = x_2^\dagger x_2$. Then there exists a sequence of matrices U_1, \dots, U_h , each of which is H or T , such that $v = U_1 \cdots U_h u$ has denominator exponent $k - 1$, or equivalently, $\rho_k(v)$ is defined and reducible.*

Proof. It can be seen from Table [table 6.1 on page 155](#) that $x_1^\dagger x_1$ is either 0000, 1010, or 0001.

- Case 1: $x_1^\dagger x_1 = x_2^\dagger x_2 = 0000$. In this case, $\rho_k(u)$ is already reducible, and there is nothing to show.
- Case 2: $x_1^\dagger x_1 = x_2^\dagger x_2 = 1010$. In this case, we know from Table [table 6.1 on page 155](#) that $x_1, x_2 \in \{0011, 0110, 1100, 1001\}$. In particular, x_1 is a cyclic

permutation of x_2 , say, $x_1 = \omega^m x_2$. Let $v = HT^m u$. Then

$$\begin{aligned}\rho_k(\sqrt{2}v) &= \rho_k\left(\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \omega^m \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\right) \\ &= \rho_k\begin{pmatrix} u_1 + \omega^m u_2 \\ u_1 - \omega^m u_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 + \omega^m x_2 \\ x_1 - \omega^m x_2 \end{pmatrix} = \begin{pmatrix} 0000 \\ 0000 \end{pmatrix}.\end{aligned}$$

This shows that $\rho_k(\sqrt{2}v)$ is twice reducible; therefore, $\rho_k(v)$ is defined and reducible as claimed.

- Case 3: $x_1^\dagger x_1 = x_2^\dagger x_2 = 0001$. In this case, we know from Table [table 6.1 on page 155](#) that $x_1, x_2 \in \{0001, 0010, 0100, 1000\} \cup \{0111, 1110, 1101, 1011\}$. If both x_1, x_2 are in the first set, or both are in the second set, then x_1 and x_2 are cyclic permutations of each other, and we proceed as in case 2. The only remaining cases are that x_1 is a cyclic permutation of 0001 and x_2 is a cyclic permutation of 0111, or vice versa. But then there exists some m such that $x_1 + \omega^m x_2 = 1111$. Letting $u' = HT^m u$, we have

$$\begin{aligned}\rho_k(\sqrt{2}u') &= \rho_k\left(\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \omega^m \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\right) \\ &= \rho_k\begin{pmatrix} u_1 + \omega^m u_2 \\ u_1 - \omega^m u_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 + \omega^m x_2 \\ x_1 - \omega^m x_2 \end{pmatrix} = \begin{pmatrix} 1111 \\ 1111 \end{pmatrix}.\end{aligned}$$

Since this is reducible, u' has denominator exponent k . Let $\rho_k(u') = (y_1, y_2)$. Because $\sqrt{2}y_1 = \sqrt{2}y_2 = 1111$, we see from Table [table 6.1 on page 155](#) that $y_1, y_2 \in \{0011, 0110, 1100, 1001\}$ and $y_1^\dagger y_1 = y_2^\dagger y_2 = 1010$. Therefore, u' satisfies the condition of case 2 above. Proceeding as in case 2, we find m' such that $v = HT^{m'} u' = HT^{m'} HT^m u$ has denominator exponent $k - 1$. This finishes the proof. \square

Lemma 6.4.2 (Column lemma). *Consider a unit vector $u \in \mathbb{D}[\omega]^n$, i.e., an n -dimensional column vector of norm 1 with entries from the ring $\mathbb{D}[\omega]$. Then there exist a sequence U_1, \dots, U_h of one- and two-level unitary matrices of types X , H , T , and ω such that $U_1 \cdots U_h u = e_1$, the first standard basis vector.*

Proof. The proof is by induction on k , the least denominator exponent of u . Let $u = (u_1, \dots, u_n)^T$.

- Base case. Suppose $k = 0$. Then $u \in \mathbb{Z}[\omega]^n$. Since by assumption $\|u\|^2 = 1$, it follows by Lemma 6.2.6 on page 154 that $\|u\|_{\text{weight}}^2 = 1$. Since u_1, \dots, u_n are elements of $\mathbb{Z}[\omega]$, their weights are non-negative integers. It follows that there is precisely one j with $\|u_j\|_{\text{weight}} = 1$, and $\|u_\ell\|_{\text{weight}} = 0$ for all $\ell \neq j$. Let $u' = X_{[1,j]}u$ if $j \neq 1$, and $u' = u$ otherwise. Now u'_1 is of the form ω^{-m} , for some $m \in \{0, \dots, 7\}$, and $u'_\ell = 0$ for all $\ell \neq 1$. We have $\omega_{[1]}^m u' = e_1$, as desired.
- Induction step. Suppose $k > 0$. Let $v = \sqrt{2}^k u \in \mathbb{Z}[\omega]^n$, and let $x = \rho_k(u) = \rho(v)$. From $\|u\|^2 = 1$, it follows that $\|v\|^2 = v_1^\dagger v_1 + \dots + v_n^\dagger v_n = 2^k$. Taking residues of the last equation, we have

$$x_1^\dagger x_1 + \dots + x_n^\dagger x_n = 0000. \quad (6.28)$$

It can be seen from Table table 6.1 on page 155 that each summand $x_j^\dagger x_j$ is either 0000, 0001, or 1010. Since their sum is 0000, it follows that there is an even number of j such that $x_j^\dagger x_j = 0001$, and an even number of j such that $x_j^\dagger x_j = 1010$.

We do an inner induction on the number of irreducible components of x . If x is reducible, then u has denominator exponent $k - 1$ by Corollary 6.2.15 on page 156, and we can apply the outer induction hypothesis. Now suppose there is some j such that x_j is irreducible; then $x_j^\dagger x_j \neq 0000$ by Lemma 6.2.13 on page 155. Because of the evenness property noted above, there must exist

some $\ell \neq j$ such that $x_j^\dagger x_j = x_\ell^\dagger x_\ell$. Applying Lemma 6.4.1 on page 180 to $u' = (u_j, u_\ell)^T$, we find a sequence \vec{U} of row operations of types H and T , making $\rho_k(\vec{U}u')$ reducible. We can lift this to a two-level operation $\vec{U}_{[j,\ell]}$ acting on u ; thus $\rho_k(\vec{U}_{[j,\ell]}u)$ has fewer irreducible components than $x = \rho_k(u)$, and the inner induction hypothesis applies. \square

Lemma 6.4.3 (Matrix decomposition). *Let U be a unitary $n \times n$ -matrix with entries in $\mathbb{D}[\omega]$. Then there exists a sequence U_1, \dots, U_h of one- and two-level unitary matrices of types X, H, T , and ω such that $U = U_1 \cdots U_h$.*

Proof. Equivalently, it suffices to show that there exist one- and two-level unitary matrices V_1, \dots, V_h of types X, H, T , and ω such that $V_h \cdots V_1 U = I$. This is an easy consequence of the column lemma, exactly as in e.g. [18, Sec. 4.5.1]. Specifically, first use the column lemma to find suitable one- and two-level row operations V_1, \dots, V_{h_1} such that the leftmost column of $V_{h_1} \cdots V_1 U$ is e_1 . Because $V_{h_1} \cdots V_1 U$ is unitary, it is of the form

$$\left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & U' \end{array} \right).$$

Now recursively find row operations to reduce U' to the identity matrix. \square

Example 6.4.4. We will decompose the matrix U from Example 6.2.18. We start with the first column u of U :

$$u = \frac{1}{\sqrt{2}^3} \begin{pmatrix} -\omega^3 + \omega - 1 \\ \omega^2 + \omega \\ \omega^3 + \omega^2 \\ -1 \end{pmatrix},$$

$$\rho_3(u) = \begin{pmatrix} 1011 \\ 0110 \\ 1100 \\ 0001 \end{pmatrix}, \quad \rho_3(u_j^\dagger u_j) = \begin{pmatrix} 0001 \\ 1010 \\ 1010 \\ 0001 \end{pmatrix}.$$

Rows 2 and 3 satisfy case 2 of Lemma 6.4.1 on page 180. As they are not aligned, first apply $T_{[2,3]}^3$ and then $H_{[2,3]}$. Rows 1 and 4 satisfy case 3. Applying $H_{[1,4]}T_{[1,4]}^2$, the residues become $\rho_3(u'_1) = 0011$ and $\rho_3(u'_4) = 1001$, which requires applying $H_{[1,4]}T_{[1,4]}$. We now have

$$H_{[1,4]}T_{[1,4]}H_{[1,4]}T_{[1,4]}^2H_{[2,3]}T_{[2,3]}^3u = v = \frac{1}{\sqrt{2}^2} \begin{pmatrix} 0 \\ 0 \\ \omega^2 + \omega \\ -\omega + 1 \end{pmatrix},$$

$$\rho_2(v) = \begin{pmatrix} 0000 \\ 0000 \\ 0110 \\ 0011 \end{pmatrix}, \quad \rho_2(v_j^\dagger v_j) = \begin{pmatrix} 0000 \\ 0000 \\ 1010 \\ 1010 \end{pmatrix}.$$

Rows 3 and 4 satisfy case 2, while rows 1 and 2 are already reduced. We reduce rows 3 and 4 by applying $H_{[3,4]}T_{[3,4]}$. Continuing, the first column is completely reduced to e_1 by further applying $\omega_{[1]}^7 X_{[1,4]}H_{[3,4]}T_{[3,4]}^3$. The complete decomposition of u is therefore given by

$$W_1 = \omega_{[1]}^7 X_{[1,4]}H_{[3,4]}T_{[3,4]}^3H_{[3,4]}T_{[3,4]} \\ H_{[1,4]}T_{[1,4]}H_{[1,4]}T_{[1,4]}^2H_{[2,3]}T_{[2,3]}^3.$$

Applying this to the original matrix U , we have $W_1U =$

$$\frac{1}{\sqrt{2}^3} \begin{pmatrix} \sqrt{2}^3 & 0 & 0 & 0 \\ 0 & \omega^3 - \omega^2 + \omega + 1 & -\omega^2 - \omega - 1 & \omega^2 \\ 0 & 0 & \omega^3 + \omega^2 - \omega + 1 & \omega^3 + \omega^2 - \omega - 1 \\ 0 & \omega^3 + \omega^2 + \omega + 1 & \omega^2 & \omega^3 - \omega^2 + 1 \end{pmatrix}.$$

Continuing with the rest of the columns, we find $W_2 = \omega_{[2]}^6 H_{[2,4]}T_{[2,4]}^3H_{[2,4]}T_{[2,4]}$, $W_3 = \omega_{[3]}^4 H_{[3,4]}T_{[3,4]}^3H_{[3,4]}$, and $W_4 = \omega_{[4]}^5$. We then have $U = W_1^\dagger W_2^\dagger W_3^\dagger W_4^\dagger$, or explicitly:

$$U = T_{[2,3]}^5 H_{[2,3]}T_{[1,4]}^6 H_{[1,4]}T_{[1,4]}^7 H_{[1,4]} \\ T_{[3,4]}^7 H_{[3,4]}T_{[3,4]}^5 H_{[3,4]}X_{[1,4]}\omega_{[1]} \\ T_{[2,4]}^7 H_{[2,4]}T_{[2,4]}^5 H_{[2,4]}\omega_{[2]}^2 H_{[3,4]}T_{[3,4]}^5 H_{[3,4]}\omega_{[3]}^4 \omega_{[4]}^3.$$

6.4.2 Main result

Consider the ring $\mathbb{Z}[\frac{1}{\sqrt{2}}, i]$, consisting of complex numbers of the form

$$\frac{1}{2^n}(a + bi + c\sqrt{2} + di\sqrt{2}),$$

where $n \in \mathbb{N}$ and $a, b, c, d \in \mathbb{Z}$. Our goal is to prove the following theorem, which was conjectured by Kliuchnikov et al. [13]:

Theorem 6.4.5. *Let U be a unitary $2^n \times 2^n$ matrix. Then the following are equivalent:*

- (a) *U can be exactly represented by a quantum circuit over the Clifford+ T gate set, possibly using some finite number of ancillas that are initialized and finalized in state $|0\rangle$.*
- (b) *The entries of U belong to the ring $\mathbb{Z}[\frac{1}{\sqrt{2}}, i]$.*

Moreover, in (a), a single ancilla is always sufficient.

Proof. First note that, since all the elementary Clifford+ T gates, as shown in ([equation \(6.27\) on page 178](#)), take their matrix entries in $\mathbb{D}[\omega] = \mathbb{Z}[\frac{1}{\sqrt{2}}, i]$, the implication (a) \implies (b) is trivial. For the converse, let U be a unitary $2^n \times 2^n$ matrix with entries from $\mathbb{D}[\omega]$. By [Lemma 6.4.3 on page 183](#), U can be decomposed into one- and two-level matrices of types X , H , T , and ω . It is well-known that each such matrix can be further decomposed into controlled-not gates and multiply-controlled X , H , T , and ω -gates, for example using Gray codes [18, Sec. 4.5.2]. But all of these gates have well-known exact representations in Clifford+ T with ancillas, see e.g. [4, Fig. 4(a) and Fig. 9] (and noting that a controlled- ω gate is the same as a T -gate). This finishes the proof of (b) \implies (a).

The final claim that needs to be proved is that a circuit for U can always be found using at most one ancilla. It is already known that for $n > 1$, an ancilla is sometimes necessary [13]. To show that a single ancilla is sufficient, in light of the above decomposition, it is enough to show that the following can be implemented with one ancilla:

- (a) a multiply-controlled X -gate;
- (b) a multiply-controlled H -gate;
- (c) a multiply-controlled T -gate.

We first recall from [4, Fig. 4(a)] that a singly-controlled Hadamard gate can be decomposed into Clifford+ T gates with no ancillas:

$$\text{---} \begin{array}{c} \bullet \\ \text{---} \end{array} \text{---} \boxed{H} \text{---} = \text{---} \boxed{S} \text{---} \boxed{H} \text{---} \boxed{T} \text{---} \oplus \text{---} \boxed{T^\dagger} \text{---} \boxed{H} \text{---} \boxed{S^\dagger} \text{---}.$$

We also recall that an n -fold controlled iX -gate can be represented using $O(n)$ Clifford+ T gates with no ancillas. Namely, for $n = 1$, we have

$$\text{---} \begin{array}{c} \bullet \\ \text{---} \end{array} \text{---} \boxed{iX} \text{---} = \text{---} \boxed{S} \text{---} \begin{array}{c} \bullet \\ \text{---} \end{array} \oplus \text{---},$$

and for $n \geq 2$, we can use

$$\text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} \boxed{iX} \text{---} = \text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} \boxed{H} \text{---} \boxed{T^\dagger} \oplus \text{---} \boxed{T} \oplus \text{---} \boxed{T^\dagger} \oplus \text{---} \boxed{T} \oplus \text{---} \boxed{H} \text{---},$$

with further decompositions of the multiply-controlled not-gates as in [5, Lem. 7.2] and [18, Fig. 4.9]. We then obtain the following representations for (a)–(c), using only one ancilla:

$$\begin{aligned} (a) \quad \text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} \boxed{X} \text{---} &= \text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} 0 \text{---} \boxed{iX} \text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} \boxed{iX} \text{---} 0 \text{---} \boxed{X} \text{---} \\ (b) \quad \text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} \boxed{H} \text{---} &= \text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} 0 \text{---} \boxed{iX} \text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} \boxed{iX} \text{---} 0 \text{---} \boxed{H} \text{---} \\ (c) \quad \text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} \boxed{T} \text{---} &= \text{---} \begin{array}{c} \bullet \\ \vdots \\ \bullet \\ \text{---} \end{array} \text{---} 0 \text{---} \boxed{iX} \text{---} \boxed{T} \text{---} \boxed{iX} \text{---} 0 \text{---}. \end{aligned}$$

□

Remark 6.4.6. The fact that one ancilla is always sufficient in Theorem 6.4.5 on the preceding page is primarily of theoretical interest. In practice, one may assume that on most

quantum computing architectures, ancillas are relatively cheap. Moreover, the use of additional ancillas can significantly reduce the size and depth of the generated circuits (see e.g. [25]).

6.4.3 The no-ancilla case

Lemma 6.4.7. *Under the hypotheses of Theorem 6.4.5 on page 185, assume that $\det U = 1$. Then U can be exactly represented by a Clifford+ T circuit with no ancillas.*

Proof. This requires only minor modifications to the proof of Theorem 6.4.5 on page 185. First observe that when an operator of the form HT^m was used in the proof of Lemma 6.4.1 on page 180, we can instead use $T^{-m}(iH)T^m$ without altering the rest of the argument. In the base case of Lemma 6.4.2 on page 182, the operator $X_{[1,j]}$ can be replaced by $iX_{[1,j]}$. Also, in the base case of Lemma 6.4.2 on page 182, whenever $n \geq 2$, the operator $\omega_{[1]}$ can be replaced by $W_{[1,2]}$, where

$$W = \begin{pmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{pmatrix}.$$

Therefore, the decomposition of Lemma 6.4.3 on page 183 can be performed so as to yield only two-level matrices of types

$$iX, \quad T^{-m}(iH)T^m, \quad \text{and } W, \tag{6.29}$$

plus at most one one-level matrix of type ω^m . But since all two-level matrices of types (equation (6.29)), as well as U itself, have determinant 1, it follows that $\omega^m = 1$. We finish the proof by observing that the multiply-controlled operators of types (equation (6.29)) possess ancilla-free Clifford+ T representations, with the latter two given by

$$\begin{array}{c} \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} = \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \\ \boxed{T^{-m}(iH)T^m} \quad \boxed{T^m} \boxed{S} \boxed{H} \boxed{T} \boxed{iX} \boxed{T^\dagger} \boxed{H} \boxed{S^\dagger} \boxed{T^{-m}} \end{array}$$

$$\begin{array}{c} \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} = \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array} \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \\ \boxed{W} \quad \boxed{iX} \boxed{T} \boxed{iX} \boxed{T^\dagger} \end{array}$$

□

As a corollary, we obtain a characterization of the n -qubit Clifford+ T group (with no ancillas) for all n :

Corollary 6.4.8. *Let U be a unitary $2^n \times 2^n$ matrix. Then the following are equivalent:*

(a) *U can be exactly represented by a quantum circuit over the Clifford+ T gate set on n qubits with no ancillas.*

(b) *The entries of U belong to the ring $\mathbb{Z}[\frac{1}{\sqrt{2}}, i]$, and:*

- $\det U = 1$, if $n \geq 4$;
- $\det U \in \{-1, 1\}$, if $n = 3$;
- $\det U \in \{i, -1, -i, 1\}$, if $n = 2$;
- $\det U \in \{\omega, i, \omega^3, -1, \omega^5, -i, \omega^7, 1\}$, if $n \leq 1$.

Proof. For (a) \implies (b), it suffices to note that each of the generators of the Clifford+ T group, regarded as an operation on n qubits, satisfies the conditions in (b). For (b) \implies (a), let us define for convenience $d_0 = d_1 = \omega$, $d_2 = i$, $d_3 = -1$, and $d_n = 1$ for $n \geq 4$. First note that for all n , the Clifford+ T group on n qubits (without ancillas) contains an element D_n whose determinant is d_n , namely $D_n = I$ for $n \geq 4$, $D_3 = T \otimes I \otimes I$, $D_2 = T \otimes I$, $D_1 = T$, and $D_0 = \omega$. Now consider some U satisfying (b). By assumption, $\det U = d_n^m$ for some m . Let $U' = U D_n^{-m}$, then $\det U' = 1$. By Lemma 6.4.7 on the previous page, U' , and therefore U , is in the Clifford+ T group with no ancillas. \square

Remark 6.4.9. Note that the last condition in Corollary 6.4.8, namely that $\det U$ is a power of ω for $n \leq 1$, is of course redundant, as this already follows from $\det U \in \mathbb{Z}[\frac{1}{\sqrt{2}}, i]$ and $|\det U| = 1$. We stated the condition for consistency with the case $n \geq 2$.

Remark 6.4.10. The situation of Theorem 6.4.5 on page 185 and Corollary 6.4.8 is analogous to the case of classical reversible circuits. It is well-known that the not-gate, controlled-not gate, and Toffoli gate generate all classical reversible functions on $n \leq 3$ bits. For $n \geq 4$

bits, they generate exactly those reversible boolean functions that define an *even permutation* of their inputs (or equivalently, those that have determinant 1 when viewed in matrix form) [17]; the addition of a single ancilla suffices to recover all boolean functions.

6.4.4 Complexity

The proof of Theorem 6.4.5 on page 185 immediately yields an algorithm, albeit not a very efficient one, for synthesizing a Clifford+ T circuit with ancillas from a given operator U . We estimate the size of the generated circuits.

We first estimate the number of (one- and two-level) operations generated by the matrix decomposition of Lemma 6.4.3 on page 183. The row operation from Lemma 6.4.1 on page 180 requires only a constant number of operations. Reducing a single n -dimensional column from denominator exponent k to $k - 1$, as in the induction step of Lemma 6.4.2 on page 182, requires $O(n)$ operations; therefore, the number of operations required to reduce the column completely is $O(nk)$.

Now consider applying Lemma 6.4.3 to an $n \times n$ -matrix with least denominator exponent k . Reducing the first column requires $O(nk)$ operations, but unfortunately, it may *increase* the least denominator exponent of the rest of the matrix, in the worst case, to $3k$. Namely, each row operation of Lemma 6.4.1 potentially increases the denominator exponent by 2, and any given row may be subject to up to k row operations, resulting in a worst-case increase of its denominator exponent from k to $3k$ during the reduction of the first column. It follows that reducing the second column requires up to $O(3(n - 1)k)$ operations, reducing the third column requires up to $O(9(n - 2)k)$ operations, and so on. Using the identity $\sum_{j=0}^{n-1} 3^j(n - j) = (3^{n+1} - 2n - 3)/4$, this results in a total of $O(3^n k)$ one- and two-level operations for Lemma 6.4.3.

In the context of Theorem 6.4.5 on page 185, we are dealing with n qubits, i.e., a $2^n \times 2^n$ -operator, which therefore decomposes into $O(3^{2^n} k)$ two-level operations. Using one ancilla, each two-level operation can be decomposed into $O(n)$ Clifford+ T gates, resulting in a total

gate count of $O(3^{2^n}nk)$ elementary Clifford+ T gates.

Chapter 7

Conclusions and future work

Bibliography

- [1] S. Abramsky. A structural approach to reversible computation. *Theoretical Computer Science*, 347(3):441–464, 2005.
- [2] S. Abramsky and B. Coecke. Physical traces: Quantum vs. classical information processing. *Electr. Notes Theor. Comput. Sci*, 69, 2002.
- [3] S. Abramsky and B. Coecke. Abstract physical traces. *Theory and Applications of Categories*, 14:114–124, 2005.
- [4] Matthew Amy, Dmitri Maslov, Michele Mosca, and Martin Roetteler. A meet-in-the-middle algorithm for fast synthesis of depth-optimal quantum circuits. Version 2, arXiv:1206.0758v2, August 2012.
- [5] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Physical Review A*, 52:3457–3467, 1995. Available from arXiv:quant-ph/9503016v1.
- [6] Erik Barendsen, Inge Bethke, Jan Heering, Richard Kennaway, Paul Klint, Vincent van Oostrom, Femke van Raamsdonk, Fer-Jan de Vries, and Hans Zantema. Cambridge University Press, The Edinburgh Building, Cambridge, CB2 2RU, UK, 2003.
- [7] Charles H. Bennet. Logical reversibility of computation. *IBM Journal of Research and Development*, 6:525–532, 1973.
- [8] Alex Bocharov and Krysta M. Svore. Resource-optimal single-qubit quantum circuits. *Physical Review Letters*, 109:190501 (5 pages), 2012. Also available from arXiv:1206.3223.
- [9] J.R.B. Cockett, Xiuzhan Guo, and Pieter Hofstra. Range categories ii: Towards regularity. Submitted for Publication, June 2012.
- [10] David Deutsch. Quantum computational networks. *Proceedings of the Royal Society of London Ser. A*, A425:73–90, 1989.
- [11] Xiuzhan Guo. *Products, Joins, Meets, and Ranges in Restriction Categories*. PhD thesis, University of Calgary, April 2012.
- [12] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to automata theory, languages, and computation*. Pearson/Addison Wesley, 3rd edition, 2007.
- [13] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. Fast and efficient exact synthesis of single qubit unitaries generated by Clifford and T gates. arXiv:1206.5236v2, June 2012.

- [14] Joachim Kock. *Frobenius Algebras and 2D Topological Quantum Field Theories*. Number 59 in London Mathematical Society Student Texts. Cambridge University Press, 2004.
- [15] Saunders Mac Lane. *Categories for the Working Mathematician*. Springer Verlag, Berlin, Heidelberg, Germany, second edition, 1997. ISBN 0-387-98403-8. Dewey QA169.M33 1998.
- [16] Ken Matsumoto and Kazuyuki Amano. Representation of quantum circuits with clifford and $\pi/8$ gates. arXiv:0806.3834v1, June 2008.
- [17] Julien Musset. Générateurs et relations pour les circuits booléens réversibles. Technical Report 97-32, Institut de Mathématiques de Luminy, 1997. Available from <http://iml.univ-mrs.fr/editions/>.
- [18] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 2RU, UK, 2000. ISBN 0 521 63235 8.
- [19] Robin Cockett. Category theory for computer science. Available at <http://pages.cpsc.ucalgary.ca/~robin/class/617/notes.pdf>, October 2009.
- [20] Robin Cockett and Stephen Lack. Restriction categories I: categories of partial maps. *Theoretical Computer Science*, 270:223–259, 2002.
- [21] Robin Cockett and Stephen Lack. Restriction categories II: Partial map classification. *Theoretical Computer Science*, 294:61–102, 2003.
- [22] Robin Cockett and Stephen Lack. Restriction categories III: colimits, partial limits, and extensivity. *Mathematical Structures in Computer Science*, 17(4):775–817, 2007. Available at <http://au.arxiv.org/abs/math/0610500v1>.
- [23] Peter Selinger. Towards a quantum programming language. *Mathematical Structures in Computer Science*, 14(4):527–586, 2004.
- [24] Peter Selinger. Dagger compact closed categories and completely positive maps. In *Proceedings of the 3rd International Workshop on Quantum Programming Languages, Chicago*, 2005.
- [25] Peter Selinger. Quantum circuits of T -depth one. *Physical Review A*, 2013. To appear. Available from arXiv:1210.0974.