Dissertation presented for the degree of Master of Science

Department of Mathematics and Applied Mathematics

University of Cape Town

On the local and global properties of information manifolds

Author:
Tslil Clingman

Supervisors:
Dr. J. Shock
Assoc. Prof. J. Murugan

August 2015

Plagiarism Declaration

I know the meaning of plagiarism and declare that all of the work in the dissertation, save for that which is properly acknowledged, is my own.

Acknowledgements

I am sincerely indebted to my supervisors, Dr. Shock and A/Prof. Murugan, for their unwavering and consistent emphasis on the entire project being, first and foremost, an enriching and rewarding learning experience on my part. Their focus on education and passion therefor, their generosity and their endless willingness to accommodate have had a strong and lasting impact upon my education, and ultimately too upon my future both within and beyond the context of academics. Without their input, guidance and desire to see others follow their passions, this project could not have been realised.

Moreover, their kindness and generosity in offering me this incredible opportunity to learn invaluable knowledge and explore an interesting field provided for the ideal way to spend this crucial period. For this alone I am deeply grateful.

Contents

1. Introduction **1**
1.1. The Fisher information metric . . . 2
1.1.1. Associated geometries . . . 2
1.1.2. Some examples . . . 2

2. Reversing the Fisher information metric **4**
2.1. Introduction . . . 4
2.2. Select constructions . . . 4
2.2.1. The $n-$dimensional sphere, $\mathbb{S}^n$ . . . 4
2.2.2. The Gaussian construction . . . 6
2.2.3. The hyperbolic secant construction . . . 7
2.3. General results . . . 8
2.4. Discussion . . . 14

3. Towards homological algebra **15**
3.1. Overview . . . 15
3.2. Monoidal categories . . . 17
3.2.1. Basic notions . . . 17
3.2.2. Braiding and symmetry . . . 23
3.2.3. Closed monoidal categories . . . 27
3.2.4. Monoids in monoidal categories . . . 32
3.3. Enriched categories . . . 37
3.3.1. Basic notions . . . 37
3.3.2. Semi-additive categories . . . 40
3.3.3. Kernels and co. . . . 47
3.3.4. Abelian categories . . . 50
3.4. Homology . . . 59
3.4.1. Exactness . . . 59
3.4.2. Chain homology . . . 62
3.4.3. The simplex category . . . 66
3.5. Discussion . . . 70

A. Appendix **71**
A.1. Adjunctions . . . 71
A.2. Regular categories . . . 75

I. Introduction

Information geometry may be described as the study of the differential structures that arise on spaces of statistical distributions, specifically those spaces whose points comprise a 'sufficiently smooth', parameterised family of probability density functions. The Fisher information metric, a notion which arises in a meaningful manner out of the information theoretic Kullback-Liebler divergence ([KL51]), provides for a way to reach one such structure by establishing a means to associate a metric tensor to such families. It thus lends itself to making definite a notion of distance between probability density functions – a concept of immediate and apparent utility in questions regarding mutual information and even gradient descent [Ama98].

Information geometry was first and principally expounded upon by S-I. Amari, the foundations of which may be found in [AN00]. Though much progress has been made in this field, we wish to draw attention to an important distinction about the very nature of the inquiry made thus far.

Questions of a somewhat local nature, 'differential' type properties, have received substantial attention – certainly the initial treatment of the field of information geometry was conducted largely from the perspective of the inherent properties of the differential structures accompanying the Fisher information metric and associated notions. The state of the art is well understood, and certain questions admit immediate and tangible answers. In particular, given a parameterised family of probability density functions the associated Fisher information metric may be stated as a concrete integral involving the family and its derivatives. However, comparatively little has been mentioned in the literature about the 'reverse direction': given a desired metric tensor, does there exist a family of probability density functions whose Fisher information metric is the desired metric tensor? Furthermore, if this assignment can be made, is it unique and what properties may we expect from the family thus recovered? The second chapter of this document is dedicated to the development of a novel answer.

Questions of a more global nature, however, have received comparatively little consideration. There exist, of course, a wealth of mathematical theories dealing with the global structure – algebrao-geometric, differential-geometric, and topological approaches for example. Of all of these, homological algebra will be explored as a vehicle to discuss such global properties and elucidate the nature of information manifolds from perhaps a different direction. The third chapter of this document deals with some of the background for, and elementary notions within, homological algebra through the indispensable language of category theory.

1. The Fisher information metric

1.1. Associated geometries

For the purposes of this work, we will assume a narrow definition of a family of probability density functions. That is, when we write 'family of probability density functions' we will mean a family of continuous functions $P_\theta : X \to \mathbb{R}$ for some domain $X \subset \mathbb{R}^n$, parameterised over $\theta \in M \subset \mathbb{R}^m$ (ie. an m-parameter family of distributions). Coordinatizing X by $x = (x^1, \ldots, x^n)$ and the parameter space M by $\theta = (\theta^1, \ldots, \theta^m)$, we will also further require that $\partial_a P_\theta := \dfrac{\partial P}{\partial \theta^a}$ is continuous on X for all $\theta \in M$. Furthermore, we will also require that every member of the family be normalised, that is,

$$(\forall \theta \in M) \int_X P(x;\theta)\,\mathrm{d}x = 1.$$

All of this may be succinctly restated as $\{P_\theta\}$ being a parametrised family of normalised, continuous functions which changes 'smoothly' over parameter space. Finally, we will refer to X as the *spatial* domain and M as the *parametric* domain, and conventionally associate the spatial domain X_i to probability density function P_i.

We now define the Fisher Information metric tensor on a finite dimensional statistical manifold. Given such a manifold, $\mathcal{M}$, whose points form a family of probability density functions with the properties listed above, there exists a Riemannian metric tensor on $\mathcal{M}$, viz.,

$$g_{ab}(\theta) = \int_X P(x;\theta)\,\partial_a \ln P(x;\theta)\,\partial_b \ln P(x;\theta)\,\mathrm{d}x. \tag{1.1.1.1}$$

The central question addressed in this paper may thus be stated as: given a Riemannian metric tensor g, under what circumstances can a family of probability density functions P be found such that the Fisher information metric tensor of P is g.

1.2. Some examples

In order to build some intuition for the relationship between a family of probability density functions and their associated metrics, we give here two examples of the computation of the Fisher metric.

Univariate Normal Distribution

Here the family of probability density functions is given by

$$P(x;\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

The distribution is parameterised by μ and σ, which we will collectively denote θ. Put another way, the manifold coordinates are given by $\theta = (\mu, \sigma)$, and the random variable is $x \in \mathbb{R}$. Note that the parametric domain is $\mathbb{R} \times \mathbb{R}^{>0}$. In order to compute g_{ab} we must compute $\partial_a \ln P$

$$\ln P = -\left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 + \ln\sigma + \ln\sqrt{2\pi}\right],$$
$$\frac{\partial}{\partial\mu}\ln P = \frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right), \quad \frac{\partial}{\partial\sigma}\ln P = \frac{1}{\sigma}\left[\left(\frac{x-\mu}{\sigma}\right)^2 - 1\right].$$

Then, using Equation 1.1.1.1, the Fisher metric for the univariate normal distribution has

$$[g] = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix} \implies \mathrm{d}s^2 = \frac{\mathrm{d}\mu^2 + 2\mathrm{d}\sigma^2}{\sigma^2}.$$

Thus we see that the Fisher metric, in this case, describes the metric tensor of a two-dimensional hyperbolic geometry. The structure on this geometry can be intuitively understood by the properties of normal distributions. In particular, for distributions with $\sigma \gg 1$, the associated 'difference' between two distributions with means μ_1 and μ_2 is less pronounced – they are harder to distinguish. For two sharply peaked distributions ($\sigma \ll 1$) with even similar μ, the difference will be very pronounced and so they are easy to distinguish. Hence the hyperbolic nature of the space.

Cauchy Distribution

The family of probability density functions for this distribution is given by

$$P(x; x_0, \gamma) = \frac{1}{\pi}\left[\frac{\gamma}{\gamma^2 + (x - x_0)^2}\right].$$

Thus, the parameter space for this family is spanned by the parameters $\theta = (x_0, \gamma) \in \mathbb{R} \times \mathbb{R}^{>0}$ and the calculation of the logarithmic derivatives gives

$$\ln P = \ln\gamma - \ln\left[\gamma^2 + (x-x_0)^2\right] - \ln\pi,$$
$$\frac{\partial}{\partial x_0}\ln P = \frac{2(x-x_0)}{\gamma^2 + (x-x_0)^2}, \quad \frac{\partial}{\partial\gamma}\ln P = \frac{1}{\gamma} - \frac{2\gamma}{\gamma^2 + (x-x_0)^2}.$$

As such, it is a simple matter to verify that the Fisher metric for the Cauchy distribution is given by

$$g_{ab} = \frac{\delta_{ab}}{2\gamma^2} \implies \mathrm{d}s^2 = \frac{1}{2}\left(\frac{\mathrm{d}x_0^2 + \mathrm{d}\gamma^2}{\gamma^2}\right).$$

The reader may wish to note that while we started with a very different distribution, the geometric structure described by its Fisher metric is very close to that of the normal distribution. In this sense, hyperbolic spaces (or Euclidean anti de-Sitter spaces) appear ubiquitous in an information geometric context.

II. Reversing the Fisher information metric

1. Introduction

Now that we have a grasp of how the Fisher information metric may be calculated and what it might represent, it would seem a natural, though likely intractable, question to ask what can be 'reached' by this computation. In particular, should we wish to prove properties about the resulting Riemannian metric tensors, it is important that we have some understanding of what this class comprises. Consider that, in the most extreme case, if all Riemannian metric tensors arise as the Fisher information metric of some family of probability density functions, there there should rightly be very little that we may say about the results in full generality.

However, it is not clear, at first glance, that it is at all possible to reverse the process of computing the Fisher metric in any meaningful way as the computation involves a definite integral of multiple powers of the underlying family of probability density functions. We examine, below, a motivating example to suggest that under certain, constrained situations such a process is indeed possible. As a prototype for a more general construction, we demonstrate how to encode the metric tensor of $\mathbb{S}^n$, for any $n \in \mathbb{N}$, in a family of one-dimensional probability density functions.

2. Select constructions

2.1. The $n-$dimensional sphere, $\mathbb{S}^n$

We begin our exploration of reversing the Fisher information computation with a one-dimensional family of probability density functions. In particular, we leverage the properties of orthonormal functions to produce a family of probability density functions which, with an appropriate set of functions h^i, give rise to the metric tensor of S^n.

Remark (II) 2.1.1. Note that for our purposes a family of univariate, real-valued functions $\{f_i(x)\}_{i\in I}$ is said to be orthonormal with weight w over a domain X if

$$\int_X f_i(x) f_j(x) w(x) \mathrm{d}x = \delta_{ij}$$

Prop. (II) 2.1.2. *Let $M \subset \mathbb{R}^n$ and $h^i \in C^1(M)$ such that*[1] $(\forall \theta \in M)\, h^i h^j \delta_{ij} = 4$ *and $\{f_i(x)\}_1^n$ be a set of orthonormal, real-valued functions with positive semidefinite weight $w(x)$ over $X \subset \mathbb{R}$. Then the family of probability density functions*

$$P(x;\theta) = \frac{1}{4}\left(\sum_{i=1}^{n} h^i(\theta) f_i(x)\right)^2 w(x), \tag{2.2.1.1}$$

gives the Fisher information metric tensor $g_{ab} = (\partial_a h^i)(\partial_b h^j)\delta_{ij}$.

Proof. That P is normalised follows from the orthonormality of f_i.

$$\frac{1}{4}\int_X \left(\sum_{i=1}^{n} h^i(\theta) f_i(x)\right)^2 w(x)\,\mathrm{d}x = \frac{1}{4}\int_X \left(\sum_{i=1}^{n}\sum_{j=1}^{n} h^i h^j f_i f_j\right) w\,\mathrm{d}x$$
$$= \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_X h^i h^j f_i f_j w\,\mathrm{d}x = \frac{1}{4} h^i h^j \delta_{ij} = 1.$$

A straightforward computation gives the desired result.

$$\begin{aligned} g_{ab} &= \int_X P(\partial_a \ln P)(\partial_b \ln P)\mathrm{d}x \\ &= \int_X w\left(\sum_{i=1}^{n} h^i f_i\right)^2 \left(\frac{\sum_{i=1}^{n}(\partial_a h^i) f_i}{\sum_{i=1}^{n} h^i f_i}\right)\left(\frac{\sum_{i=1}^{n}(\partial_b h^i) f_i}{\sum_{i=1}^{n} h^i f_i}\right)\mathrm{d}x \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\int_X (\partial_a h^i)(\partial_b h^j) f_i f_j w\,\mathrm{d}x = (\partial_a h^i)(\partial_b h^i)\delta_{ij}. \end{aligned}$$

■

We pause to note that we may view the above statement, $g_{ab} = (\partial_a h^i)(\partial_b h^i)\delta_{ij}$, as the result of applying the transition functions h to the flat Euclidean metric δ. As such, and noting that we required $h^i h^j \delta_{ij} = 4$, we immediately infer that

Cor. (II) 2.1.3. *The metric tensor of $\mathbb{S}^n$ can be reached as the Fisher Information metric of the distribution eq.* (2.2.1.1) *where h is the transition function from $\mathbb{E}^n$ to $4\mathbb{S}^n$, the n-dimensional sphere of radius four.*

In the above we have shown a general way to find a given metric tensor in terms of the transition functions from flat Euclidean space to a desired geometry. However, there is a specific condition on the h^i given by $h^i h_i = 4$ which constrains these strongly. In what follows, we will generalise this result in a way which will remove this constraint.

[1] Here we use Einstein summation and the lowered and raised indices have no differential geometric interpretation other than to aid in the appropriate summations.

2.2. The Gaussian construction

Now that we have reason to believe that it is possible, at least in special cases, to pick a metric tensor and construct a family of probability density functions whose Fisher information metric is the selected metric, we attempt to extend our results to arbitrary Riemannian metrics.

Consider a family of probability density functions given by a product of n, uncorrelated, disjoint, one-dimensional Gaussian probability density functions with unit variance. Explicitly,

$$P(x;\theta) = \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(x^i - h^i(\theta)\right)^2\right), \tag{2.2.2.1}$$

where M, the parametric domain, is not yet fixed, $X = \mathbb{R}^n$, and $h^i \in C^1(M)$. From this, we may compute the Fisher information metric as follows

$$\begin{aligned}
g_{ab} &= \frac{1}{\sqrt{(2\pi)^n}} \int_X \mathrm{d}x\, e^{-\frac{1}{2}\sum_{i=1}^{n}\left(x^i-h^i\right)^2} \left(\sum_{j=1}^{n}(\partial_a h^j)\left(x^j - h^j\right)\right)\left(\sum_{k=1}^{n}(\partial_b h^k)\left(x^k - h^k\right)\right) \\
&= \frac{1}{\sqrt{(2\pi)^n}} \int_X \mathrm{d}x\, e^{-\frac{1}{2}\sum_{i=1}^{n}\left(x^i-h^i\right)^2} \left(\sum_{j=1}^{n}(\partial_a h^j)(\partial_b h^j)\left(x^j - h^j\right)^2 + \begin{matrix}\text{vanishing}\\ \text{cross-terms}\end{matrix}\right) \\
&= \sum_i \left\{(\partial_a h^i)(\partial_b h^i)\prod_k\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{d}x^k\, e^{-\frac{1}{2}\left(x^i-h^i\right)^2}\left(x^k - h^k\right)^2\right)\right\}.
\end{aligned}$$

It is a simple matter to complete the computation to obtain

$$g_{ab} = (\partial_a h^j)(\partial_b h^k)\delta_{jk}. \tag{2.2.2.2}$$

This result allows us enough flexibility to be able to always give an h and M such that g_{ab} may be constructed as desired. In particular, we may begin at eq. (2.2.2.2) and read backwards to find eq. (2.2.2.1). In doing so, we fix a desired g_{ab} and accompanying manifold $\mathcal{M}$, and attempt to realise an h and M for which eq. (2.2.2.2) would hold. Unlike the case of prop. (II) 2.1.2, which came with the constraint $h^i h_i = 4$, this process is here always possible.

The Nash Embedding Theorem [Nas56] tells us that there is an $n \in \mathbb{N}$ such that $(\mathcal{M}, g)$ may be C^1 isometrically embedded in $(\mathbb{E}^n, \delta)$. Specifically then, it tells us that there exists an h such that $g = h^*\delta$. As such, interpreting eq. (2.2.2.2) as the statement that g is the pullback of δ via h we see that we need only select an n large enough to accommodate the Nash embedding of the desired manifold $\mathcal{M}$ in $\mathbb{E}^n$ (which is always possible) and we have h and M to satisfy the arrangement. Consequently, we have a family of probability density functions, given by eq. (2.2.2.1) whose Fisher information metric is the desired, arbitrary Riemannian metric.

Said another way, eq. (2.2.2.2) states simply that g_{ab} is the pullback from a higher dimensional flat space to a manifold embedded in that space, via h. In the case of coincidence of dimensions between g and h, the result bears the simple interpretation of h acting as a set of transition functions from δ to g.

The metric of $\mathbb{S}^2$

To cement the understanding of the importance and generality of eq. (2.2.2.2) we construct the metric tensor of $\mathbb{S}^2$. Suppose we desire a family of probability density functions whose Fisher information metric is the metric tensor of $\mathbb{S}^2$. Specifically, if the unit sphere has line element

$$ds^2 = \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2,$$

then we can proceed as outlined above, and write down a set of transition functions

$$h = (\cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi),$$

from $\mathbb{E}^3$ to the embedded $\mathbb{S}^2$. Applying the construction of eq. (2.2.2.1) we find

$$P(x, y, z; \theta, \phi) = (2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}\left((x-\cos\theta\sin\phi)^2 + (y-\sin\theta\sin\phi)^2 + (z-\cos\phi)^2\right)}.$$

This is easily recognisable as a product of three Gaussian probability density functions, each with a mean which is periodic in the parameters. This means that we have the geometry and topology of a sphere, where each point on the sphere corresponds to a three dimensional Gaussian distribution with unit variance and mean denoted by the point on the sphere. This exercise can be performed for any $\mathbb{S}^n$ by simply forming the appropriate h.

The ease with which we are able to perform this construction is indicative of the power underlying eq. (2.2.2.2) and the accompanying statement that *any Riemannian metric tensor may be reached via this construction*.

2.3. The hyperbolic secant construction

In the previous subsection we gave a construction based upon a product of Gaussian probability density functions and demonstrated its flexibility. Now we demonstrate that the above-mentioned results are just as achievable with an entirely different family of probability density functions. Consider the family

$$P = \frac{1}{\pi^n} \prod_{i=1}^{n} \operatorname{sech}\left(x^i - h^i\sqrt{2}\right).$$

Other than the functional dependence on $\sim x^i - h^i$, this is entirely different from the Gaussians discussed earlier. However, computing the Fisher information metric we find the result to be of that most general form

$$g_{ab} = (\partial_a h^i)(\partial_b h^j)\delta_{ij}.$$

Naturally, this bears the same interpretation as the previous result and serves to suggest that relatively little of the information about the original family of probability density functions is carried through to the metric tensor itself.

The careful reader will note that we now have *two* means to the same end, and may wonder just how many more ways we may achieve the above result. Indeed the following section serves to introduce a general framework which will show that the answer is that there is an 'infinite-fold' degeneracy in the construction, and thus there is always an 'infinite-to-one' mapping between families of PDFs and Riemannian metrics via the Fisher information metric.

3. General results

In this section we will elaborate on a more general set of statements which allow for definitions independent of dimensionality and functional dependence of the parameters of the PDF in question. We begin by showing how to construct a family of spatially disjoint probability density functions out of individual families of probability density functions.

Def. (II) 3.0.1. The spatially disjoint product of two families of probability density functions on the same parametric domain, $P_1 = P_1(x^1, \ldots, x^k; \theta) : X_1 \times M \to \mathbb{R}$ and $P_2 = P_2(x^1, \ldots, x^n; \theta) : X_2 \times M \to \mathbb{R}$, is defined as

$$(P_1 \odot P_2)(x^1 \ldots, x^{n+k}; \theta) = P_1(x^1, \ldots, x^n; \theta) \cdot P_2(x^{n+1}, \ldots, x^{n+k}; \theta).$$

Note that $P_1 \odot P_2 : (X_1 \times X_2) \times M \to \mathbb{R}$ and we write $P^{\odot n}$ where we mean $\bigodot_{i=1}^{n} P$.

Given this, we will here show how a special property of spatially disjoint products underpins all the general results achieved in this work. That is, the Fisher information metric transforms the spatially disjoint product of probability density functions into a sum of their corresponding, individually considered metric tensors.

Thm. (II) 3.0.2. *If $P = P(x; \theta)$ is a probability density function with a decomposition $P = \bigodot P_i^{\odot e_i}$ for some P_i and $e_i \in \mathbb{N}^+$ then $g_{ab}\left(\bigodot P_i^{\odot e_i}\right) = \sum e_i g_{ab}(P_i)$.*

Proof. Let us rewrite $P = \bigodot \hat{P}_i^{\odot e_i} = \bigodot P_j$ where each $\hat{P}_i$ has been accumulated into the spatially disjoint product e_i times, that is, $P_j = \hat{P}_i$ for e_i many j. Then, in order to compute $g(P)$ we expand logarithmic derivatives to arrive at

$$g_{ab}(P) = \sum_i \sum_j \int_X \mathrm{d}x \, \frac{P}{P_i P_j} (\partial_a P_i)(\partial_b P_j).$$

To proceed we must evaluate the double sum, and to do so we examine the cases $j = i$ and $j \neq i$ separately. In the event of the latter, $j \neq i$, we have

$$\int_X \mathrm{d}x \, \frac{P}{P_i P_j} (\partial_a P_i)(\partial_b P_j) = \left(\int_{X_i} \mathrm{d}x^a \cdots \mathrm{d}x^k \, \partial_a P_i \right) \left(\int_{X_j} \mathrm{d}x^m \cdots \mathrm{d}x^r \, \partial_b P_j \right),$$

where we have expanded the integral as a product over its disjoint spatial domains and have suppressed all other terms as they were of the form $\int_{X_i} \mathrm{d}x^a \cdots \mathrm{d}x^k P_i = 1$. Moreover, we note that P_i satisfies the conditions (by the definition of the probability density function) for the exchange of integral and derivative and so

$$\int_{X_i} \mathrm{d}x^a \cdots \mathrm{d}x^k \partial_a P_i = \partial_a \int_{X_i} \mathrm{d}x^a \cdots \mathrm{d}x^k P_i = \partial_a(1) = 0.$$

Thus contributions from terms where $j \neq i$ is zero. On the other hand, the cases for which $i = j$ admit simple resolution as

$$\int_X \mathrm{d}x \frac{P}{P_i P_i} (\partial_a P_i)(\partial_b P_i) = \int_{X_i} \mathrm{d}x^a \cdots \mathrm{d}x^k (\partial_a P_i)(\partial_b P_i) \frac{1}{P_i} = g_{ab}(P_i),$$

where again we have expanded the integral as a product and suppressed all terms whose integral was one. Finally, we recall that we had exactly e_i many P_j such that $P_j = \hat{P}_i$ and so we collect e_i many such contributions of $g_{ab}(P_j)$. ∎

Remark (II) 3.0.3. That we essentially require $M_1 = M_2 = M$ in the definition of the spatially disjoint product is a matter of some subtlety. Consider that if $M_1 \neq M_2$ we would be within reason to set $M = M_1 \times M_2$ and reinterpret the definition as

$$(P_1 \odot P_2)(x^1 \ldots, x^{n+k}; \theta, \phi) = P_1(x^1, \ldots, x^n; \theta) \cdot P_2(x^{n+1}, \ldots, x^{n+k}; \phi).$$

In this case, however, $g(P)$ is not strictly the sum of $g(P_i)$ as the latter may all be of different dimension. Simply re-interpreting P_i to have enlarged parametric domain M will not solve this problem as then it may happen that $g(P_i)$ will no longer be non-degenerate and so not a metric tensor. Thus, the direct ability of the above result to "glue" together disjoint metric tensors is apparent, but nuanced and not an immediate consequence of the exposition given.

In effect then, care should be taken when examining the statement $g(\bigodot P_i) = \sum g(P_i)$ so as to ensure that it is done with the understanding that $g(P_i)$ is to have zero entries where appropriate for the purpose of the sum, but not when considered as its own metric tensor. More formally, we could write $g(\bigodot P_i) = \sum \tilde{g}(P_i)$ where $\tilde{g}$ is expressed precisely as g, but is extended to all of M as suggested above, and is free from interpretation as a metric tensor. Hereafter, it is taken for granted that such nuances are appreciated by the reader.

The importance of thm. (II) 3.0.2 cannot be overstated. From here on, it is simply a matter of finding convenient forms of $g_{ab}(P_i)$ for some parameterisation of P_i so that we may take $\bigodot P_i$ and arrive at a desired metric tensor. That is, if we can find a P_i such that $g_{ab}(P_i) \propto (\partial_a h^i)(\partial_b h^i)$ then we can take $P = \bigodot P_i$ to find $g_{ab} \propto (\partial_a h^i)(\partial_b h^j)\delta_{ij}$ by the above. Here, the whole is more than the sum of its parts – given $g_{ab} \propto (\partial_a h^i)(\partial_b h^j)\delta_{ij}$ we are able to find an h for our desired manifold and then create a desired P out of constituent P_i, each containing some part of $\{h^i\}$. Beginning with disjoint P_i, however, the qualities which the individual distributions should exhibit, to attain a given g, are not clear. Furthermore, we note here that while $\bigodot P_i$ will yield the desired result, if we find multiple families of probability density functions, we may equally well combine them to achieve the same result.

Thus, what we really seek are simple forms of functional dependence of families of probability density functions upon our set of differentiable functions h so that explicit computations may be made. Recall that we saw, in the calculations in sections 2.2.2 and 2.2.3, that we may leverage reparameterisation invariance of spatial domains to our advantage. Such symmetries of the spatial domain allow us to essentially eliminate any functional dependence of the integrals upon the h^i and produce multiplicative factors of $\partial_a h$ in the process. To that end, we explore a generalisation of the symmetry used in the above-mentioned subsections.

Prop. (II) 3.0.4. *Fix a one-dimensional probability density function $\hat{P}(x)$ on X for which X remains invariant under the change of variables $y = f(x;\theta)$, for some differentiable family of diffeomorphisms $f : X \times M \to X$ (the parameter space is M) and let $P(x;\theta) = f_x(x;\theta)\hat{P}(f(x;\theta))$ such that $\partial_a P \not\equiv 0$ where we write f_x for $\frac{\partial f}{\partial x}$ and f_a for $\partial_a f$. Then*

$$g_{ab}(P) = \int_X \frac{f_{ax}f_{bx}}{(f_x)^2}\hat{P}(y) + \left(\frac{\partial(f_a f_b)}{\partial y} + f_a f_b \frac{d\ln\hat{P}(y)}{dy}\right)\frac{d\hat{P}(y)}{dy}dy, \tag{4.4}$$

where we assume that we have written all functions in terms of $y = f(x;\theta)$ using the expression $x = f^{-1}(y;\theta)$ where necessary.

Proof. We first check that $P(x;\theta) = f_x(x;\theta)\hat{P}(f(x;\theta))$ is normalised. To that end, let $y = f(x;\theta)$

$$\int_X P\mathrm{d}x = \int_X f_x\hat{P}\mathrm{d}x = \int_X f_x\hat{P}\frac{\mathrm{d}y}{f_x} = 1.$$

Then we compute the logarithmic derivatives necessary for the Fisher information metric

$$\partial_a \ln P = \frac{1}{f_x\hat{P}(f)}\left(\frac{\mathrm{d}\hat{P}(f)}{\mathrm{d}f}(f_a f_x) + \hat{P}(f)(f_{ax})\right).$$

We proceed with the computation by making the change of variables $y = f(x;\theta)$

$$\begin{aligned} g_{ab} &= \int_X \frac{1}{(f_x)^2\hat{P}(f)}\left(\frac{\mathrm{d}\hat{P}(y)}{\mathrm{d}y}(f_a f_x) + \hat{P}(y)(f_{ax})\right)\left(\frac{\mathrm{d}\hat{P}(y)}{\mathrm{d}y}(f_b f_x) + \hat{P}(y)(f_{bx})\right)\mathrm{d}y \\ &= \int_X f_a f_b \frac{\mathrm{d}P(y)}{\mathrm{d}y}\frac{\mathrm{d}\ln P(y)}{\mathrm{d}y} + \frac{f_{ax}f_{bx}}{(f_x)^2}P(y) + \left(\frac{f_a f_{bx} + f_b f_{ax}}{f_x}\right)\frac{\mathrm{d}P(y)}{\mathrm{d}y}\mathrm{d}y. \end{aligned}$$

Finally, we recognise that $\frac{\partial}{\partial x} = f_x\frac{\partial}{\partial y}$ and that $f_a f_{bx} + f_b f_{ax} = \frac{\partial(f_a f_b)}{\partial x}$, and collect terms to arrive at the result. ■

Of course, examining symmetry at such an abstract level cannot be expected to yield concrete answers immediately and so that the statement of prop. (II) 3.0.4 is opaque and not obviously useful is not surprising. Indeed, in what follows we make various simplifying assumptions about the functional form of the symmetry function f to arrive at generalisations of familiar results.

We begin by noticing that there is a term in eq. (4.4) which is proportional to $f_a f_b$. If it could be arranged that $f_a f_b$ be independent of y, then we could simply extract a term proportional to $f_a f_b$ from the result – a term whose importance we already know. Moreover, if we could ensure that the other terms vanish, we would have $g_{ab} \propto f_a f_b$ and achieve our general result once more.

To that end, we choose to require that f_x be constant and $f_{ax} = 0$. Although this is likely not the *only* way to achieve our desired effect, it will certainly suffice. In this case, we see immediately that $f(x;\theta) = cx + h(\theta)$ is the general solution – but this is nothing other than the statement of translation invariance. Thus, we may achieve the following results by means of prop. (II) 3.0.4.

Prop. (II) 3.0.5. *Fix a one-dimensional probability density function* $\hat{P}$ *such that the change of variables* $y = x - h$ *for* $h(\theta)$ *a differentiable function on* $M \subset \mathbb{R}^m$ *leaves the spatial domain* X *unchanged. Let* $P(x;\theta) = \hat{P}\,(x - h)$ *then* $g_{ab} = (\partial_a h)(\partial_b h)D$ *where*

$$D = \int_X dx \left(\frac{\partial P(x)}{\partial x}\right)\left(\frac{\partial \ln P(x)}{\partial x}\right).$$

Proof. Apply prop. (II) 3.0.4 to $f(x;\theta) = x - h(\theta)$. ∎

Cor. (II) 3.0.6. *Fix one-dimensional probability density functions* P_i *and let* $h^i(\theta)$ *be differentiable on* $M \subset \mathbb{R}^m$ *and write* $y^i = x^i - h^i$ *such that* X_i *is unchanged under this change of variables for all* i. $P(x;\theta) = \bigodot P_i\left(x^i - h^i\right)^{\odot e_i}$ *gives* $g_{ab}(P) = (\partial_a h^i)(\partial_b h^j)D_{ij}$ *where*

$$D_{ij} = \begin{cases} e_i \int_{X_i} dx^i \left(\frac{\partial P_i}{\partial x^i}\right)\left(\frac{\partial \ln P_i}{\partial x^i}\right), & i = j \\ 0, & i \neq j \end{cases}$$

Proof. Combine prop. (II) 3.0.5 and thm. (II) 3.0.2. ∎

Remark (II) 3.0.7. When P_i are all Gaussian, $D_{ij} = \delta_{ij}$ and so the result of eq. (2.2.2.2) follows as a special case.

To demonstrate how one might achieve the encoding of an arbitrary Riemannian metric tensor into a spatially disjoint product of one-dimensional families of probability density functions, consider the following example.

Example (II) 3.0.8

Suppose we desire a hyperbolic metric tensor g whose associated line element is given by $\frac{1}{\beta}(\mathrm{d}\alpha^2 + \mathrm{d}\beta^2)$, on the open subset $M = \{(\alpha,\beta) \in \mathbb{R}^2 \mid \beta > 1\} \subset \mathbb{H}^2$. With some work, it can be shown that an isometric embedding of M into $\mathbb{R}^3$ can be achieved through the function

$$h = \left(\frac{\cos\alpha}{\beta}, \frac{\sin\alpha}{\beta}, \ln\left(\beta + \sqrt{\beta^2 - 1}\right) - \frac{\sqrt{\beta^2-1}}{\beta} \right).$$

That is, $g = h^*\delta$. Moreover, it is evident that h is at least C^1 so we may apply our construction to it and write, for example,

$$P = P_1\left(x - h^1\right) \odot P_2\left(y - h^2\right) \odot P_3\left(z - h^3\right),$$

for any one-dimensional probability density functions P_i which satisfy translation invariance as outlined in prop. (II) 3.0.5. By cor. (II) 3.0.6 we then know that $g(P) = h^*D$ and so the result follows in the case that $D = \delta$.

In particular then, we may choose to let $X_i = \mathbb{R}$ for $i \in \{1,2,3\}$ and put

$$\hat{P}_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \hat{P}_2(x) = \frac{1}{\pi}\operatorname{sech} x, \quad \hat{P}_3(x) = \frac{1}{\pi\left(1 + x^2\right)},$$

for which $D_1 = 1$ and $D_2 = D_3 = \frac{1}{2}$. Thus, taking the values of D_i into account, we may write $P(x,y,z;\alpha,\beta) = \hat{P}_1\left(x - h^1\right) \odot \hat{P}_2\left(y - \sqrt{2}h^2\right) \odot \hat{P}_3\left(z - \sqrt{2}h^3\right)$ to recover

$$P(x,y,z;\alpha,\beta) = \frac{\left(\sqrt{2\pi^5}\right)^{-1} \operatorname{sech}\left(x - \frac{\sqrt{2}\sin\alpha}{\beta}\right) e^{-\frac{1}{2}\left(y - \frac{\cos\alpha}{\beta}\right)^2}}{1 + \left[z + \frac{\sqrt{2\beta^2-2}}{\beta} - \sqrt{2}\ln\left(\beta + \sqrt{\beta^2-1}\right)\right]^2},$$

defined on $\mathbb{R}^3 \times M$, and for which we know, due to cor. (II) 3.0.6, the metric tensor is $g = \beta^{-2}\delta$. It may also be verified directly that, given,

$$P_1(x;\alpha,\beta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(x - \frac{\cos\alpha}{\beta}\right)^2}, \quad P_2(x;\alpha,\beta) = \frac{1}{\pi}\operatorname{sech}\left(x - \frac{\sqrt{2}\sin\alpha}{\beta}\right),$$

$$P_3(x;\alpha,\beta) = \frac{\pi^{-1}}{1 + \left[x + \frac{\sqrt{2\beta^2-2}}{\beta} - \sqrt{2}\ln\left(\beta + \sqrt{\beta^2-1}\right)\right]^2}, \text{ we have}$$

$$g(P_1) = \frac{1}{\beta^4}\begin{bmatrix} \beta^2\sin^2\alpha & \beta\sin\alpha\cos\alpha \\ \beta\sin\alpha\cos\alpha & \cos^2\alpha \end{bmatrix}, \quad g(P_2) = \frac{1}{\beta^4}\begin{bmatrix} \beta^2\cos^2\alpha & -\beta\sin\alpha\cos\alpha \\ -\beta\sin\alpha\cos\alpha & \sin^2\alpha \end{bmatrix},$$

$$g(P_3) = \frac{1}{\beta^4}\begin{bmatrix} 0 & 0 \\ 0 & \beta^2-1 \end{bmatrix},$$

whose sum is as desired – that is, $g\left(\bigodot P_i\right) = \sum g(P_i)$ as thm. (II) 3.0.2 assured us. Thus, we have managed to encode a desired metric tensor as the Fisher information metric of a spatially disjoint product of three, one-dimensional families of probability density functions.

We can explore another possible simplifying form of transformation f. Consider that were $f(x;\theta) \propto x$, then every term in eq. (4.4) would contribute a factor proportional to f_a. Again, this is a desirable result and so we explore the symmetry of scale invariance.

Prop. (II) 3.0.9. *Fix a one-dimensional probability density function $\hat{P}$ such that the change of variables $y = xe^h$ for $h(\theta)$ a differentiable function on $M \subset \mathbb{R}^m$ leaves the spatial domain X unchanged. Let $P(x;\theta) = e^h \hat{P}\left(xe^h\right)$ then $g_{ab} = (\partial_a h)(\partial_b h)E$ where*

$$E = \int\limits_X P(x)\left(1 + x\frac{\partial \ln P(x)}{\partial x}\right)^2 dx.$$

Proof. We set $f(x;\theta) = e^{h(\theta)}x$ and compute the required derivatives for prop. (II) 3.0.4 as follows

$$f_a = \partial_a h x e^h, \quad f_x = e^h, \quad f_{ax} = \partial_a h e^h, \quad \frac{\partial(f_a f_b)}{\partial y} = 2(\partial_a h)(\partial_b h)y.$$

The result follows straightforwardly. ■

Cor. (II) 3.0.10. *Fix one-dimensional probability density functions P_i and let $h^i(\theta)$ be differentiable on $M \subset \mathbb{R}^m$ and write $y^i = x^i e^{h^i}$ such that X_i is unchanged under this change of variables for all i. $P(x;\theta) = \bigodot e^{h^i} P_i\left(x^i e^{h^i}\right)^{\odot e_i}$ gives $g_{ab}(P) = (\partial_a h^i)(\partial_b h^j)E_{ij}$ where*

$$E_{ij} = \begin{cases} e_i \int_{X_i} dx^i P^i \left(1 + x\frac{\partial \ln P_i}{\partial x^i}\right)^2, & i = j \\ 0, & i \neq j \end{cases}$$

Proof. Combine prop. (II) 3.0.9 and thm. (II) 3.0.2. ■

Cor. (II) 3.0.11. *Every Riemannian metric tensor may be reached as the result of the Fisher information metric acting upon a spatially disjoint product of families of one-dimensional probability density functions.*

Proof. Apply either cor. (II) 3.0.10 or cor. (II) 3.0.6 to the desired C^1 pullback h, which exists due to the isometric embedding of the desired manifold in $\mathbb{E}^n$ via the Nash Embedding theorem. ■

It can now be seen that relatively simple computations give rise to highly useful results by way of thm. (II) 3.0.2. Indeed, to extend this work one need only find other families of probability density functions whose Fisher information metric can be made to be proportional to $(\partial_a h)(\partial_b h)$ in order to combine them in the requisite multiplicity to allow h to be the pullback for a desired Riemannian metric tensor. That we made explicit use of spatial domain symmetries using prop. (II) 3.0.4 should be seen as merely a convenient and intuitive way of making use of thm. (II) 3.0.2 to construct desirable results.

4. Discussion

We have thus seen that we are able to associate a Riemannian information manifold with a concretely-stated metric tensor to a given family of probability density functions. However, such an association immediately suggests the consideration of the possibility of reversing the process and so asks just how much information about the underlying family of PDFs is preserved in the computation of the Fisher information metric.

By observing that the Fisher information metric takes a spatially disjoint product of families of probability density functions to a sum of their individual Fisher information metrics, we were able to give a positive existence result to the guiding question and even an explicit construction when a suitable pullback is known. That is, every Riemannian metric tensor is realisable as the Fisher information metric of some family of probability density functions exhibiting a select symmetry on their spatial domains – cor. (II) 3.0.11. The symmetry is crucial to our construction and is the tool that effectively enables an injection of dependence upon the components of the isometric embedding used. Furthermore, what we have shown is that very little information is carried through the association, so little indeed, that any family of PDFs exhibiting the necessary properties will give rise to the chosen Riemannian metric.

III. Towards homological algebra

1. Overview

Homological algebra may be described as that branch of mathematics which concerns itself with the study of homology in a generalised sense. Though the present context constrains our interest in this area to its ability to elucidate certain geometric properties of manifolds, we shall nevertheless enjoy a winding tour through various supporting ideas so that when we are finally in a position to begin to consider its applications, we may do so with confidence and with the ability to transfer our knowledge to various related areas. Chief among all themes in this chapter will be that of taking as general a standpoint as we dare, all but necessitating the use of category theory.

The histories of homological algebra and category theory are strongly intertwined and it is not the place of this document, nor indeed the author, to attempt to discern cause from effect. A comprehensive motivation for the use of category theory in homological algebra would necessitate going into, in the words of [McL90], “the array of homology theories at the time and the forefront of 1940s abstract algebra, and we would do this without using category theory, and we would waste a lot of time on things category theory has now made much easier.” Furthermore, understanding category theory or homological algebra in this way would suffer from a paradoxical inability to give motivating examples. This is because “precisely the examples serious enough to have motivated the definitions are too hard to be worth giving now without the benefit of categorical hindsight.”

Thus we shall visit various notions and topics within category theory without feeling the insatiable need for historical motivation or contextualisation, instead being confident that our direction is justified by the very existence of modern homological algebra – itself almost certainly only tractable within the language of categories.

The general course we shall be charting through the surrounding theory will follow, by and large, the idea of introducing additional structure to categories – either through positing the existence of certain objects or, later, by requiring morphism collections to exhibit structures all their own – until the categories we consider demonstrate a sufficiently rich theory to both support and generalise select properties, facets of whose interplay enable a discussion of homological concepts. That is to say, we shall first establish a working understanding of monoidal categories which will enable us to discuss enriched categories which, in turn, will set the stage for abelian categories whose properties support generalised notions of our eventual destinations, viz., homology and exactness.

Regrettably, any course through such a rich and general theory will inevitably err through omission. This is particularly the case in this work, as its time-frame demanded of the author a certain singular focus upon only one destination. Many interesting and undoubtedly important avenues and considerations thus remain unexplored – this work is far from comprehensive, even when restricted to those concepts covered, and is not intended to be otherwise. Our aim in exploring the abovementioned theory is to prepare ourselves for more directed research and inquiry. Once we have concluded our exposition, we will be in a position to begin to contemplate the many ways in which homological algebra may be applied to the specific case of information geometry in the hopes of descrying non-trivial properties of the objects of central concern – information manifolds.

2. Monoidal categories

A simple starting point in endowing a category with extra structure would be to somehow induce an algebraic structure upon its objects. Of the various algebraic structures at hand, monoids stand out as demanding relatively little in the way of structure, but offering a rich enough theory to be of interest. As such, we will attempt to view $\operatorname{Obj}\mathfrak{C}$ as an algebraic monoid.

2.1. Basic notions

Def. (III) 2.1.1. A monoidal category $\mathfrak{C}$ is a category equipped with

1. a bifunctor $\otimes : \mathfrak{C} \times \mathfrak{C} \to \mathfrak{C}$, the tensor product
2. an object $I \in \operatorname{Obj}\mathfrak{C}$, the identity object
3. a natural isomorphism $\alpha_{A,B,C} : (A \otimes B) \otimes C \to A \otimes (B \otimes C)$, the associator
4. a natural isomorphism $\lambda : I \otimes \bullet \to \mathrm{id}_{\mathfrak{C}}$, the left unitor
5. a natural isomorphism $\rho : \bullet \otimes I \to \mathrm{id}_{\mathfrak{C}}$, the right unitor

such that the following diagrams commute for all $A, B, C, D \in \operatorname{Obj}\mathfrak{C}$. The monoidal category is called strict when α, λ, ρ are identities.

$$
\begin{array}{ccc}
(A \otimes I) \otimes B & \xrightarrow{\alpha_{A,I,B}} & A \otimes (I \otimes B) \\
& {\scriptstyle \rho_A \otimes \mathrm{id}_B} \searrow \quad \swarrow {\scriptstyle \mathrm{id}_A \otimes \lambda_B} & \\
& A \otimes B &
\end{array}
$$

$$
\begin{array}{ccccc}
& & ((A \otimes B) \otimes C) \otimes D & & \\
& {\scriptstyle \alpha_{A,B,C} \otimes \mathrm{id}_D} \swarrow & & \searrow {\scriptstyle \alpha_{A \otimes B,C,D}} & \\
(A \otimes (B \otimes C)) \otimes D & & & & (A \otimes B) \otimes (C \otimes D) \\
\downarrow {\scriptstyle \alpha_{A,B \otimes C,D}} & & & & \downarrow {\scriptstyle \alpha_{A,B,C \otimes D}} \\
A \otimes ((B \otimes C) \otimes D) & & \xrightarrow[\mathrm{id}_A \otimes \alpha_{B,C,D}]{} & & A \otimes (B \otimes (C \otimes D))
\end{array}
$$

Note that we do not require the components of the unitors and associator to be identities, only isomorphisms. Moreover, naturality ensures that they commute with arrows in the category – if $f : A \to B$ then $\lambda_B(\mathrm{id}_I \otimes f) = f\lambda_A$ as arrows from $I \otimes A$ to B, for instance – thereby maximally preserving the monoidal structure.

Remark (III) 2.1.2. If we truly wish to express the difference between the monoidal category $(\mathfrak{C}, \otimes, I, \alpha, \lambda, \rho)$ and the 'underlying' category $\mathfrak{C}$, we will write $\mathfrak{C}_0$ for the latter. In this way, a functor $F : \mathfrak{C}_0 \to \mathfrak{D}_0$ is an 'ordinary' functor.

Another notational convention is the omission of the $\otimes$ symbol between objects in favour of juxtaposition where unambiguous, and the writing of $\mathrm{id}_A \otimes f$ and $f \otimes \mathrm{id}_A$ as Af and fA respectively. This convention has the convenient side effect that some identities are obvious ($\mathrm{id}_A \otimes \mathrm{id}_B = \mathrm{id}_{A\otimes B}$ just reads $AB = AB$).

Before we divert the reader's attention with examples and surrounding theory, there is an immediate question which may seem, at first glance, somewhat troublesome. Notice that both λ_I and ρ_I are natural isomorphisms $II \to I$. We do not explicitly require these to coincide, and it may seem strange to leave such matters up to chance. To quell such unsettling ideas, we give the following results.

Prop. (III) 2.1.3. *In any monoidal category, the following hold for all objects A and B:*

1. $\lambda_{IA} = I\lambda_A$ *and* $\rho_{AI} = \rho_A I$
2. $\lambda_{AB}\alpha_{I,A,B} = \lambda_A B$ *and* $\rho_{AB} = (A\rho_B)\alpha_{A,B,I}$
3. $\lambda_I = \rho_I$

Proof. The first statement follows from the naturality squares for λ and ρ. Observe that $\lambda_A\lambda_{IA} = \lambda_A(I\lambda_A)$ and similarly for ρ, and as λ, ρ are isomorphisms we have (1).

With this established, we turn to the proof of (2). Though perhaps long, it is entirely mechanical relying only on the naturality of α and the pentagonal and triangular identities in def. (III) 2.1.1. For these reasons, we explicitly only show the first statement, as the second is entirely similar.

To begin, observe that if we wish to show $f = g$ for generic arrows $f, g : A \rightrightarrows B$, it suffices to show that $If = Ig$ as then $\lambda_B(If) = \lambda_B(Ig)$ but by naturality $\lambda_B(If) = f\lambda_A$ and $f\lambda_A = g\lambda_A \implies f = g$. [1]

With that established, we will show that $I(\lambda_{AB}\alpha_{I,A,B}) = I(\lambda_A B)$.

$$\begin{aligned}
&I(\lambda_{AB}\alpha_{I,A,B}) = (I\lambda_{AB})(I\alpha_{I,A,B}) \\
&= (I\lambda_{AB})\alpha_{I,I,AB}\alpha^{-1}{}_{I,I,AB}(I\alpha_{I,A,B}) && \text{(triangular identity)} \\
&= (\rho_I(AB))\alpha^{-1}{}_{I,I,AB}(I\alpha_{I,A,B}) \\
&= \alpha_{I,A,B}\alpha^{-1}{}_{I,A,B}(\rho_I(AB))\alpha^{-1}{}_{I,I,AB}(I\alpha_{I,A,B}) \\
&= \alpha_{I,A,B}((\rho_I A)B)\alpha^{-1}{}_{II,A,B}\alpha^{-1}{}_{I,I,AB}(I\alpha_{I,A,B}) && \text{(naturality of } \alpha^{-1}) \\
&= \alpha_{I,A,B}((I\lambda_A)\alpha_{I,I,A}B)\alpha^{-1}{}_{II,A,B}\alpha^{-1}{}_{I,I,AB}(I\alpha_{I,A,B}) && \text{(triangular identity)} \\
&= \alpha_{I,A,B}((I\lambda_A)B)(\alpha_{I,I,A}B)\alpha^{-1}{}_{II,A,B}\alpha^{-1}{}_{I,I,AB}(I\alpha_{I,A,B}) \\
&= (I(\lambda_A B))\alpha_{I,IA,B}(\alpha_{I,I,A}B)\alpha^{-1}{}_{II,A,B}\alpha^{-1}{}_{I,I,AB}(I\alpha_{I,A,B}) && \text{(naturality of } \alpha) \\
&= I(\lambda_A B) && \text{(pentagonal identity)}
\end{aligned}$$

Finally, we note that the equality $\lambda_I I \overset{(2)}{=} \lambda_{II}\alpha_{I,I,I} \overset{(1)}{=} (I\lambda_I)\alpha_{I,I,I} \overset{\text{(tri.)}}{=} \rho_I I$ is sufficient to give (3), and thereby conclude the proof. ■

[1] That is, λ and ρ give an equivalence of categories where the functors are $I\otimes$ and $\otimes I$ respectively.

With that established, we divert our attention to examples of monoidal categories. Perhaps the most obvious starting point is the interpretation of a commutative monoid as a category.

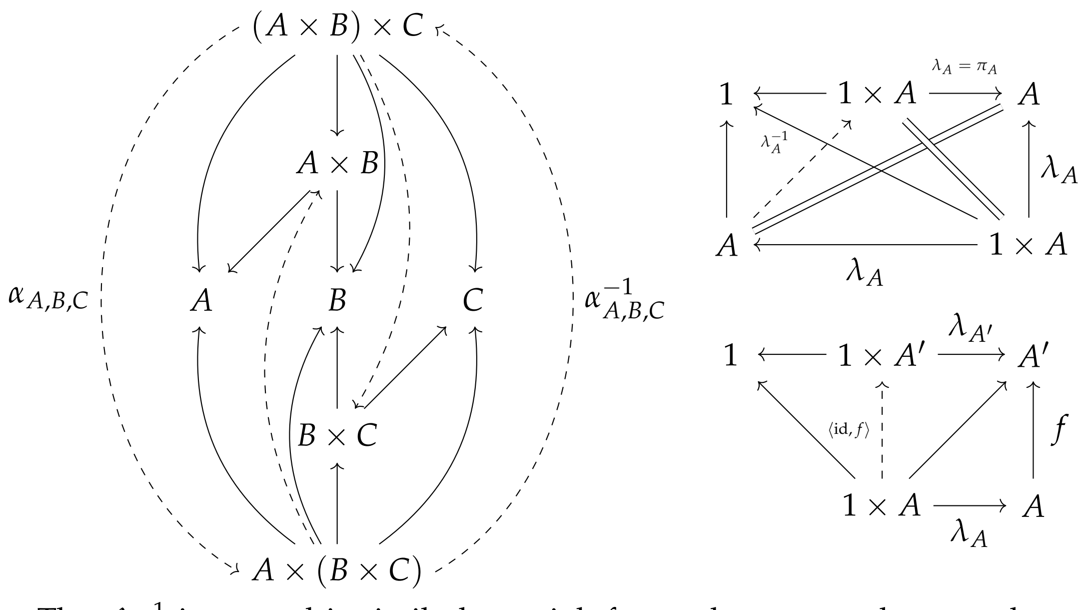
Example (III) 2.1.4

Given a commutative algebraic monoid, $(M, \cdot, e)$, we can create a category $\mathfrak{C}$ with $\operatorname{Obj}\mathfrak{C} = M$ and $\mathfrak{C}(a,b) = \{m \in M \,|\, ma = b\}$ where composition is then just multiplication. Now, we may endow $\mathfrak{C}$ with a strict monoidal structure as follows. We define the tensor product to be the underlying monoidal operation on both morphisms and objects. The tensor product is indeed a bifunctor, if we have $f : a \to a' \iff fa = a'$ and $g : b \to b' \iff gb = b'$ then we certainly have $fgab = fagb = a'b'$ and so $fg : ab \to a'b'$. To complete the construction, we take the identity object to be e, and define $\alpha_{a,b,c} = \mathrm{id}_{abc} = e$ as the monoidal operation is associative, and $\lambda_a = \rho_a = \mathrm{id}_a = e$. That these maps are natural follows from the fact that they are identities.

In the event that the algebraic monoid is not commutative, the above construction will not work as the tensor product cannot be extended to morphisms. In such situations, and indeed more generally, we may instead turn to the following construction.

Example (III) 2.1.5

A cartesian monoidal category is a category $\mathfrak{C}$ which supports all finite products, endowed with the monoidal structure of $\otimes = \times$, $I = 1$ and $\alpha_{A,B,C}, \lambda_A, \rho_A$ the canonical isomorphisms arising from the below commutative diagrams.



That λ^{-1} is natural is similarly straightforward to see, and so too that ρ, ρ^{-1} are natural isomorphisms. That α is natural is also true, but not demonstrated explicitly in the above diagrams.

The reader should be quick to note that there is a dual to the above, a *cocartesian* monoidal category. If a category supports all finite *coproducts* then we may view $(\mathfrak{C}, 0, +)$ as a monoidal category with α, λ, ρ the canonical isomorphisms.

As in the above two examples, there are many situations in which the tensor product behaves like a product in some sense of the word. Perhaps an important class of such examples is when the tensor product is actually a tensor product of modules or vector spaces.

Example (III) 2.1.6

The category Ab of abelian groups admits a monoidal structure. In particular, if we view abelian groups as $\mathbb{Z}$-modules then the bifunctor can be the tensor product of $\mathbb{Z}$-modules, $\otimes_{\mathbb{Z}}$, with $\mathbb{Z}$ serving as the identity and α, λ, ρ the canonical isomorphisms. That is, we would define $\lambda_G(n \otimes g) = ng$ and $\lambda^{-1}{}_G(ng) = 1 \otimes ng$, and similarly for ρ, from which it is easy to check that λ, ρ are natural isomorphisms.

Finally, to demonstrate that tensor products needn't be products in any sense of the word, we move to examine the category of endofunctors on a category.

Example (III) 2.1.7

The category $[\mathfrak{C}, \mathfrak{C}] = \operatorname{End} \mathfrak{C}$ is monoidal with the tensor product being composition of functors, the Godement product on natural transformations, and the identity object as $\mathrm{id}_{\mathfrak{C}}$. Furthermore, $\alpha_{A,B,C} = \mathrm{id}_{ABC}$ as composition is associative, and it is readily apparent that $\lambda_A = \rho_A = \mathrm{id}_A$ and that α, β, ρ are natural. In fact, the category of endofunctors is thus a strict monoidal category under functor composition.

Prop. (III) 2.1.8. *The unit object in a monoidal category is unique up to isomorphism.*

Proof. Let $\mathfrak{C}$ be a monoidal category with unit objects I and I', and associated unitors λ, ρ and λ', ρ'. Observe that $f = \lambda_{I'}(\rho'_I)^{-1} : I \to I'$ is an isomorphism as it is the composite of two isomorphisms. ■

Given this fact, and drawing on inspiration from Set and Cat, we will begin to think of arrows from I to an object of the category C as *generalised elements* of C. Although the morphisms are not in any meaningful way 'contained in' C, they do (in many ways) speak for the way that C 'behaves'. The reader is advised to entertain this notion and be aware of how it repeatedly reappears in the context of monoidal categories (c.f. monoid objects, closed monoidal categories, *etc*).

Now that we have a working understanding of how monoidal categories may manifest themselves, and which abstractions they attempt to capture, we may ask the natural question: do they form a category? In order to answer this, we must first define functors between monoidal categories. In doing so, we notice that there are varying degrees to which we may seek to have the functor respect the monoidal nature of the categories at hand.

Def. (III) 2.1.9. Given two monoidal categories $(\mathfrak{C}, I_{\mathfrak{C}})$ and $(\mathfrak{D}, I_{\mathfrak{D}})$, a monoidal functor between them is given by a triplet (F, ϕ, ε) where $F : \mathfrak{C}_0 \to \mathfrak{D}_0$ is a functor, $\phi_{A,B} : FA\,FB \to F(AB)$ is a natural transformation and $\varepsilon : I_{\mathfrak{D}} \to FI_{\mathfrak{C}}$ is a morphism in $\mathfrak{D}$, such that for all $A, B, C \in \operatorname{Obj} C$ the following diagrams commute.

$$
\begin{array}{ccc}
(FA\,FB)FC & \xrightarrow{\alpha_{\mathfrak{D}}} & FA(FB\,FC) \\
{\scriptstyle \phi_{A,B}\,FC}\big\downarrow & & \big\downarrow{\scriptstyle FA\,\phi_{B,C}} \\
F(AB)\,FC & & FA\,F(BC) \\
{\scriptstyle \phi_{AB,C}}\big\downarrow & & \big\downarrow{\scriptstyle \phi_{A,BC}} \\
F((AB)C) & \xrightarrow[F\alpha_{\mathfrak{C}}]{} & F(A(BC))
\end{array}
$$

$$
\begin{array}{ccc}
I_{\mathfrak{D}}\,FA & \xrightarrow{\lambda_{\mathfrak{D}}} & FA \\
{\scriptstyle \varepsilon\,FA}\big\downarrow & & \big\uparrow{\scriptstyle F\lambda_{\mathfrak{C}}} \\
FI_{\mathfrak{C}}\,FA & \xrightarrow[\phi_{I_{\mathfrak{C}},A}]{} & F(I_{\mathfrak{C}}A)
\end{array}
\qquad
\begin{array}{ccc}
FA\,I_{\mathfrak{D}} & \xrightarrow{\rho_{\mathfrak{D}}} & FA \\
{\scriptstyle FA\,\varepsilon}\big\downarrow & & \big\uparrow{\scriptstyle F\rho_{\mathfrak{C}}} \\
FA\,FI_{\mathfrak{C}} & \xrightarrow[\phi_{A,I_{\mathfrak{C}}}]{} & F(AI_{\mathfrak{C}})
\end{array}
$$

The first diagram expresses how ϕ, the 'factoring' operation, and F should respect the associative nature of the monoidal categories. If both categories are strict, then the diagram reduces to a commuting square expressing simply that the order in which we 'factor' out F does not matter. The next two diagrams express for us that 'factoring' out F should respect the left and right unitors of the categories. This, the reader should bear in mind, is analogous to the idea of monoid homomorphisms for which we have $f(a)f(b)f(c) = f(abc)$, among other identities.

The reader would do well to scrutinise the previous statement. Indeed, the above diagram essentially only gives 'one direction' of the identity present for monoid homomorphisms, viz., $f(a)f(b)f(c) \to f(abc)$. In particular, no restrictions were placed on the invertibility of the 'factoring' operation or indeed the unit morphism. It is thus here that we have an opportunity to discriminate among monoidal functors, through the degree to which they are true to our analogy.

Def. (III) 2.1.10. A monoidal functor (F, ϕ, ε) is termed

- Lax, if it merely satisfies the conditions present in def. (III) 2.1.9.
- Strong, if ϕ is a natural isomorphism and ε is an isomorphism.
- Strict, if ϕ and ε are identities.

With a basic understanding of monoidal functors between monoidal categories in hand, we move to address the notion of monoidal natural transforms between monoidal functors. Intuitively, we expect that a monoidal natural transform should respect the functorial nature in the ordinary way, that it should respect 'factoring' morphism in a natural manner, and that it should take identity morphisms to identity morphisms. Indeed,

Def. (III) 2.1.11. If $F_1, F_2 : \mathfrak{C} \rightrightarrows \mathfrak{D}$ are two monoidal functors between monoidal categories then a monoidal natural transform $\tau : F_1 \to F_2$ is a natural transform of the functors $F_1, F_2 : \mathfrak{C}_0 \rightrightarrows \mathfrak{D}_0$ such that the following diagrams commute for all $A, B \in \operatorname{Obj} C$.

$$\begin{array}{ccc} F_1A\,F_1B & \xrightarrow{\phi_1} & F_1(AB) \\ {\scriptstyle \tau_A \otimes \tau_B}\downarrow & & \downarrow{\scriptstyle \tau_{AB}} \\ F_2A\,F_2B & \xrightarrow[\phi_2]{} & F_2(AB) \end{array} \qquad \begin{array}{ccc} & I & \\ {\scriptstyle \varepsilon_1}\swarrow & & \searrow{\scriptstyle \varepsilon_2} \\ F_1I & \xrightarrow[\tau_I]{} & F_2I \end{array}$$

2.2. Braiding and symmetry

Now that we have a grasp on the elementary notions concerning monoidal categories, we may be tempted to investigate certain properties of such categories. In particular, should we examine SET (or indeed any cartesian monoidal category) we notice that we have $A \times B \cong B \times A$ naturally in both arguments. However, it is clear that this is not the case in a general monoidal category. Despite being strict, $\operatorname{End} \mathfrak{C}$ does not exhibit such behaviour and thus we must introduce this at the level of a structural property.

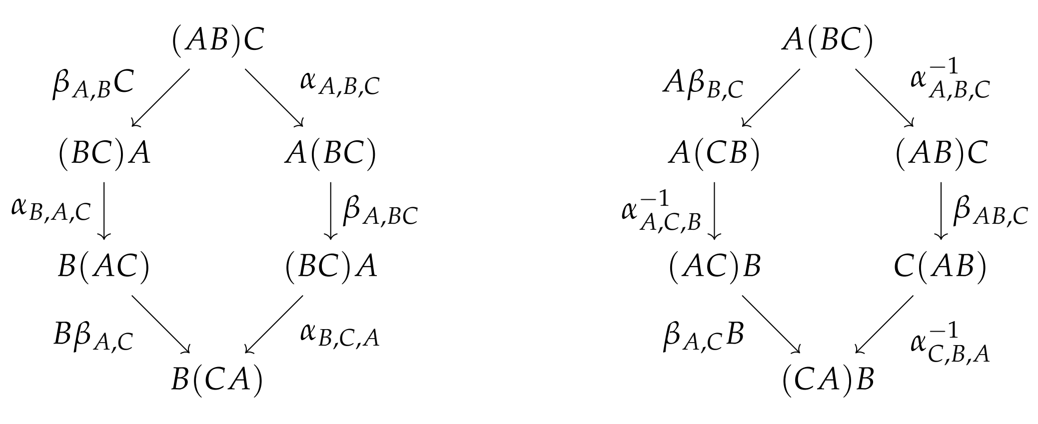
However, there is another hidden privilege that cartesian monoidal categories enjoy. If we let $\beta_{A,B} : A \times B \to B \times A$ be the natural isomorphism then it is a matter of some triviality that $\beta_{A,B}\beta_{B,A} = \mathrm{id}_{A\times B}$. However, as the example below illustrates, this is not generally the case.

Example (III) 2.2.1

> Recall that a ring R is N-graded, for a monoid N, if there exist a family of subgroups $(R_n)_{n\in N}$ such that $R = \bigoplus R_n$ and $R_n \cdot R_m \subseteq R_{n+m}$. Further, recall that if M is a right R-module, M is a graded right R-module if there exist a family of subgroups $(M_n)_{n\in N}$ such that $M = \bigoplus M_n$ and $M_m \cdot R_n \subseteq M_{m+n}$.
>
> If R is a graded commutative ring, then we may form GMod_R, the category of graded R-modules, whose objects are graded R-modules and whose morphisms are graded module morphisms. That is, if $f : M \to N$ for $M, N \in \operatorname{Obj} \mathrm{GMod}_R$ then $f(M_n) \subseteq N_n$ and f is otherwise an R-module morphism.
>
> It is a simple matter to verify that defining $(M \otimes N)_k = \sum_{m+n=k} M_m \otimes_R N_n$ endows the category with a monoidal structure. Moreover, it can be shown [JS68] that braidings for GMod_R are in bijection with invertible elements r of R, and are given by $\beta_{M,N}(a \otimes b) = r^{mn}(b \otimes a)$ for $a \in M_m$ and $b \in N_n$. It is clear that, in general, $\beta^2 \neq \mathrm{id}$.

Given this example, it is clear now that we cannot in general require that $\beta^2 = \mathrm{id}$ as it is in the cartesian monoidal case. However, the general notion stands and we define, with suitable coherence requirements,

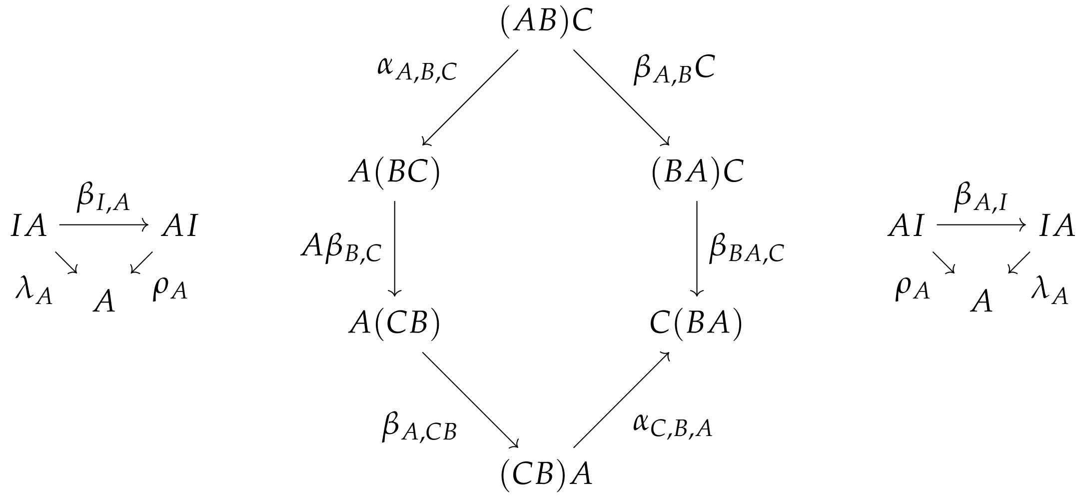
Def. (III) 2.2.2. A braided monoidal category is a monoidal category equipped with a binatural isomorphism $\beta_{A,B} : AB \to BA$ such that the following diagrams commute for all $A, B, C \in \operatorname{Obj} \mathfrak{C}$.



In the above, the second diagram can be seen to be identical to the first with $\beta_{A,B}$ replaced by ${\beta^{-1}}_{B,A}$, thereby indicating that β and β^{-1} are allowed to be somehow 'different'. Moreover, the first diagram encapsulates the notion that we may either braid A through BC in 'one-step' or successively 'pull' it through B and then C, and then reach the same result. Similarly, the second diagram says that we may braid AB through C in 'one-step' or gradually, without changing the result. In a strict monoidal category, the diagrams give $\beta_{A,BC} = (B\beta_{A,C})(\beta_{A,B}C)$ and $\beta_{AB,C} = (\beta_{A,C}B)(A\beta_{B,C})$.

As we may expect, these coherence conditions suffice to prove that the braiding respects 'reasonable' operations (in particular, unitors and associators). The following is a theorem of [JS68], given here without proof.

Thm. (III) 2.2.3 (Joyal, Street). *In a braided monoidal category, the following diagrams commute.*

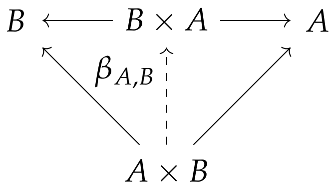


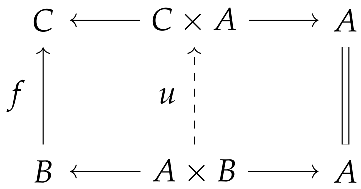
Of course, to justify our generalisation we are able to show that cartesian monoidal categories are canonically braided.

Prop. (III) 2.2.4. *Every cartesian monoidal category admits a canonical braiding.*

Proof. First we define β, then we show it is natural, and finally demonstrate that the left hexagonal diagram commutes – the proof of the right one is entirely similar.

We define β through the universal property in the below-left diagram, and then make use of the below-right diagram to show its naturality in its second argument – the case for the first is entirely similar. Let A, B, C be objects in the category, and let $f : B \to C$, then we wish to show that $(f \times \mathrm{id}_A)\beta_{A,B} = \beta_{A,C}(\mathrm{id}_A \times f)$. Both of these morphisms are arrows $A \times B \to C \times A$ and so we need only check that they have the universal properties given below-right to show equality.



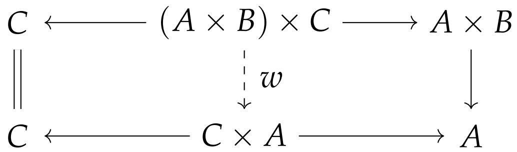


It is a simple matter to check that $\pi_C(f \times \mathrm{id}_A)\beta_{A,B} = f\pi_A = \pi_C\beta_{A,C}(\mathrm{id}_A \times f)$ and similarly that $\pi_A(f \times \mathrm{id}_A)\beta_{A,B} = \pi_B = \pi_A\beta_{A,C}(\mathrm{id}_A \times f)$, thus β is natural in its second argument.

That the hexagonal diagrams commute is nothing more than a drawn-out exercise in universal property arguments. By drawing out the large commutative diagram for $g = (\mathrm{id}_B \times \beta_{A,C})\alpha_{B,A,C}(\beta_{A,B} \times \mathrm{id}_C)$ we see that $g : (A \times B) \times C \to B \times (C \times A)$ is characterised through the universal properties $\pi_B g = \pi_B \pi_{A\times B}$ and $\pi_{C\times A} g = \beta_{A,C} v$ where $v : (A \times B) \times C \to A \times C$ is the unique map with $\pi_A v = \pi_A \pi_{A\times B}$, $\pi_C v = \pi_C$.

Should we perform a similar exercise for $h = \alpha_{B,C,A}\beta_{A,B\times C}\alpha_{A,B,C}$ we see that $\pi_B h = \pi_B \pi_{A\times B}$ as desired, but this time we have that $\pi_{C\times A} h = v'\beta_{A,B\times C}\alpha_{A,B,C}$ where $v' : (B \times C) \times A \to C \times A$ is the unique map with $\pi_A v' = \pi_A$, $\pi_C v' = \pi_C \pi_{B\times C}$.

In order to show equality between $\beta_{A,C} v$ and $v'\beta_{A,B\times C}\alpha_{A,B,C}$ as parallel morphisms $(A \times B) \times C \to C \times A$, we turn to a final universal property argument.



Using the commuting diagram above, we check that

$$
\begin{aligned}
&\pi_A\beta_{A,C}v = \pi_A v = \pi_A\pi_{A\times B}\\
&\pi_B\beta_{A,C}v = \pi_C v = \pi_C\\
&\pi_A v'\beta_{A,B\times C}\alpha_{A,B,C} = \pi_A\beta_{A,B\times C}\alpha_{A,B,C} = \pi_A\alpha_{A,B,C} = \pi_A\pi_{A\times B}\\
&\pi_C v'\beta_{A,B\times C}\alpha_{A,B,C} = \pi_C\pi_{B\times C}\beta_{A,B\times C}\alpha_{A,B,C} = \pi_C\pi_{B\times C}\alpha_{A,B,C} = \pi_C
\end{aligned}
$$

thus completing the proof. ■

As in the case of monoidal categories, we also briefly discuss the behaviour of functors which preserve the braided nature of their categories.

Def. (III) 2.2.5. A braided monoidal functor $F : \mathfrak{C} \to \mathfrak{D}$ is a monoidal functor between the monoidal categories for which the following diagram commutes for all $A, B \in \operatorname{Obj} \mathfrak{C}$.

$$\begin{array}{ccc} FA\,FB & \xrightarrow{\beta_{FA,FB}} & FB\,FA \\ \phi_{A,B}\downarrow & & \downarrow\phi_{B,A} \\ F(AB) & \xrightarrow[F\beta_{A,B}]{} & F(BA) \end{array}$$

Remark (III) 2.2.6. Natural transformations between braided monoidal functors are monoidal natural transforms between the monoidal functors, and are not required to satisfy any additional properties.

Ultimately, however, our interest in braided monoidal categories is constrained to the case of symmetry.

Def. (III) 2.2.7. A braided monoidal category is called symmetric when the braiding has $\beta^2 = \mathrm{id}$.

Remark (III) 2.2.8. Symmetric monoidal functors are simply braided monoidal functors where the braiding happens to be a symmetry.

As is no doubt already manifestly evident,

Prop. (III) 2.2.9. *Every cartesian monoidal category with the canonical braiding is symmetric.*

2.3. Closed monoidal categories

Remark (III) 2.3.1. This section will frequently make use of various properties of adjunctions. For the convenience of the reader, the requisite supporting theory has been exhibited in appendix A.1.

Not for the last time, we look to Set as a cartesian monoidal category for more interesting structural properties which we may take to the general monoidal case. The facet that catches our eye this time is that if A and B are sets, then $\textsc{Set}(A, B) \in \operatorname{Obj} \textsc{Set}$. In some ways, this is the result of a special privilege enjoyed by the somewhat central role of Set in the standard theory. However, a more careful treatment of this property is desirable.

In order to effect this, we must first essay the 'categorification' of our understanding of sets $\textsc{Set}(B, C)$. Classically, we would write such a set as C^B and we would be quick to note that we have a canonical morphism $\mathrm{ev} : C^B \times B \to C$ as defined by $\mathrm{ev}(f, b) = f(b)$. Upon careful inspection, we may surmise that C^B and ev are universal with respect to the following property.

Prop. (III) 2.3.2. *For all sets A, and set functions $f : A \times B \to C$ there exists a unique set function $\lambda f : A \to C^B$ such that* $\mathrm{ev}\langle \lambda f, \mathrm{id}_B \rangle = f$.

Proof. Given f we define $(\lambda f)(a)(b) = f(a, b)$. It is a simple matter to verify that $\mathrm{ev}\langle \lambda f, \mathrm{id}_B \rangle = f$ and uniqueness follows from pointwise agreement. ∎

Should we inspect the above with an eye to a more general theory, we note that the universal property is telling us something about the functor $\times B$. In particular, we may recognise that (C^B, ev) is somehow a universal arrow from $\times B$ to C. That is, for every $(A, f : A \times B \to C)$ we have a unique morphism $A \to C^B$ such that the appropriate diagram commutes. Moreover, we know that if every object has a universal arrow, as is the case in Set for $\times B$, then we have an adjunction! Thus, in one broad and permeating stroke we generalise as much as seems reasonable and define the following.

Def. (III) 2.3.3. A right-closed monoidal category is a monoidal category $\mathfrak{C}$ wherein for every object $C \in \operatorname{Obj} \mathfrak{C}$ the functor $\bullet \otimes C$ has a right adjoint $[C, \bullet]$. That is, for every $A, B, C \in \operatorname{Obj} \mathfrak{C}$, $\mathfrak{C}(A \otimes B, C) \cong \mathfrak{C}(A, [B, C])$ naturally in A, C. The image of the functor $[C, \bullet]$ is called the internal morphism object. Similarly, a left-closed monoidal category is one in which the functor $C \otimes \bullet$ has a right adjoint $[\bullet, C]$.

Remark (III) 2.3.4. Drawing upon the previous section, we immediately see that if the monoidal category is braided, then it is left-closed iff it is right-closed iff it is biclosed and so we simply say that it is a closed braided monoidal category. Importantly, in this case the isomorphism of (external) morphism objects is natural in *all* arguments.

To demonstrate that we have indeed generalised the classical theory of SET, and the extent to which internal morphism objects behave as though they were actually collections of morphisms, we begin with the following, perhaps presumptuous, definition.

Def. (III) 2.3.5. In a right-closed monoidal category, we

- say that a morphism $a : I \to A$ is called a point of A.
- define $\mathrm{ev}_{A,B} : [A,B] \otimes A \to A$ to be the counit of the adjunction on $\otimes A$, so named as it plays the role of internal evaluation,
- and define $\circ_{A,B,C} : [B,C] \otimes [A,B] \to [A,C]$ to be the image of the morphism $\mathrm{ev}_{B,C}(\mathrm{id}_{[B,C]} \otimes \mathrm{ev}_{A,B})\alpha_{[B,C],[A,B],A} : ([B,C] \otimes [A,B]) \otimes A \to C$ under the adjunction isomorphism $\mathfrak{C}(([B,C] \otimes [A,B]) \otimes A, C) \to \mathfrak{C}([B,C] \otimes [A,B], [A,C])$, so named as it plays the role of internal composition.

In a right-closed monoidal category, we have $\mathfrak{C}(A,B) \cong \mathfrak{C}(I \otimes A, B) \cong \mathfrak{C}(I, [A,B])$, where the first isomorphism is $\mathfrak{C}(\lambda_A, B)$, and so we can identify arrows $f : A \to B$ with points $[f] : I \to [A,B]$ of internal morphism objects. That is, morphisms from I to the internal morphism object are simply elements of the external morphism object. For the first time then, the reader may disregard the insistence of the author and see for himself that there is a sense in which we should regard morphisms from I to an object as generalised elements (*points*) of the object in question.

With this language established, in the following two propositions, we examine the interplay between external and internal morphism objects and the properties of internal composition.

Prop. (III) 2.3.6. *In a right-closed monoidal category $\mathfrak{C}$, for all objects A, B, C,*

1. *the adjoint isomorphism $\mathfrak{C}(A, [B,C]) \to \mathfrak{C}(A \otimes B, C)$ takes $f : A \to [B,C]$ to* $\mathrm{ev}_{B,C}(f \otimes \mathrm{id}_B)$
2. *for all $f : A \to B$,* $\mathrm{ev}_{A,B}([f] \otimes \mathrm{id}_A) \cong f$
3. *let $a : I \to A$ be a point, and $f : A \to B$, then* $\mathrm{ev}_{A,B}([f] \otimes a) \cong fa$*, a point of B*
4. $\mathrm{ev}_{B,C}\left(\mathrm{id}_{[B,C]} \otimes \mathrm{ev}_{A,B}\right)\alpha_{[B,C],[A,B],A} = \mathrm{ev}_{A,C}\left(\circ_{A,B,C} \otimes \mathrm{id}_A\right)$

Proof. We immediately recognise (1) as a trivial consequence of the definition of adjunctions in terms of units and counits. To elaborate the point, if we write the adjoint isomorphism as $\phi_{A,C} : \mathfrak{C}(A, [B,C]) \to \mathfrak{C}(A \otimes B, C)$ then we know (prop. (A) 1.0.14) that it can be stated in terms of the counit and left adjoint functor as $\phi_{A,C} = \mathrm{ev}_{B,C}(- \otimes \mathrm{id}_A)$.

That (2) holds follows from the fact that $\mathrm{ev}_{A,B}([f] \otimes \mathrm{id}_A) = f\lambda_A$, by the definition of $[f]$ and (1). For (3), $\mathrm{ev}_{A,B}([f] \otimes a) = \mathrm{ev}_{A,B}([f] \otimes \mathrm{id}_A)(\mathrm{id}_A \otimes a) = f\lambda_A(\mathrm{id}_I \otimes a) = fa\lambda_I$ by (2) and naturality of λ. Finally, (4) follows again through a simple combination of definition and (1). ■

Prop. (III) 2.3.7. *In a right-closed monoidal category $\mathfrak{C}$, for all objects A, B, C, D*

1. *Let $f : A \to B$ and $g : B \to C$ then $\circ_{A,B,C}([g] \otimes [f]) \cong [gf]$*
2. $\circ_{A,B,D}(\circ_{B,C,D} \otimes \mathrm{id}_{[A,B]}) = \circ_{A,C,D}(\mathrm{id}_{[C,D]} \otimes \circ_{A,B,C})\alpha_{[C,D],[B,C],[A,B]}$ *as morphisms $([C,D] \otimes [B,C]) \otimes [A,B] \to [A,D]$ – composition is associative within the monoidal structure*
3. $\circ_{A,B,B}([\mathrm{id}_B] \otimes \mathrm{id}_{[A,B]}) = \lambda_{[A,B]}$ *and* $\circ_{A,A,B}(\mathrm{id}_{[A,B]} \otimes [\mathrm{id}_A]) = \rho_{[A,B]}$ *– composition is unital within the monoidal structure*
4. $[A \otimes B, C] \cong [A, [B,C]]$

Proof. Let us begin by writing ${\phi^{-1}}_{A,C} : \mathfrak{C}(AB, C) \to \mathfrak{C}(A, [B,C])$ for the adjunction isomorphism, natural in both arguments. The statement of left-hand side of (1) then becomes ${\phi^{-1}}_{[B,C][A,B],C}(\mathrm{ev}_{B,C}(\mathrm{id}_{[B,C]} \otimes \mathrm{ev}_{A,B})\alpha_{[B,C],[A,B],A})([g] \otimes [f])$. However, ϕ^{-1} is natural and so we turn to the appropriate naturality square to proceed with simplification.

$$\begin{array}{ccc}
\mathfrak{C}(([B,C][A,B])A, C) & \xrightarrow{{\phi^{-1}}_{[B,C][A,B],C}} & \mathfrak{C}([B,C][A,B], [A,C]) \\
\Big\downarrow \mathfrak{C}(([g] \otimes [f])A, C) & & \Big\downarrow \mathfrak{C}([g] \otimes [f], C) \\
\mathfrak{C}((II)A, C) & \xrightarrow[{\phi^{-1}}_{II,C}]{} & \mathfrak{C}(II, [A,C])
\end{array}$$

Thus, we have the equalities

$$\begin{aligned}
&{\phi^{-1}}_{[B,C][A,B],C}(\mathrm{ev}_{B,C}([B,C]\,\mathrm{ev}_{A,B})\alpha_{[B,C],[A,B],A})([g] \otimes [f]) \\
=&{\phi^{-1}}_{II,C}\left(\mathrm{ev}_{B,C}([B,C]\,\mathrm{ev}_{A,B})\alpha_{[B,C],[A,B],A}\left(([g] \otimes [f])\, A\right)\right) && \text{(naturality of } \phi^{-1}\text{)} \\
=&{\phi^{-1}}_{II,C}\left(\mathrm{ev}_{B,C}([B,C]\,\mathrm{ev}_{A,B})\left([g] \otimes ([f]A)\right)\alpha_{I,I,A}\right) && \text{(naturality of } \alpha\text{)} \\
=&{\phi^{-1}}_{II,C}\left(\mathrm{ev}_{B,C}([g] \otimes f\lambda_A)\alpha_{I,I,A}\right) && \text{(prop. (III) 2.3.6 (2))} \\
=&{\phi^{-1}}_{II,C}\left(g\lambda_B(I\,(f\lambda_A))\alpha_{I,I,A}\right) && \text{(prop. (III) 2.3.6 (2))} \\
=&{\phi^{-1}}_{II,C}\left(gf\lambda_A\,(I\lambda_A)\,\alpha_{I,I,A}\right) && \text{(naturality of } \lambda\text{)} \\
=&{\phi^{-1}}_{II,C}\left(gf\lambda_A\,(\rho_I A)\right) && \text{(def. (III) 2.1.1, triangle)} \\
=&{\phi^{-1}}_{II,C}\left(gf\lambda_A\,(\lambda_I A)\right) && \text{(prop. (III) 2.1.3 (3))}
\end{aligned}$$

In order to progress from here, we must introduce an isomorhpism which will allow us to recast the above into $[gf]$. We carefully note that the naturality of ϕ^{-1} gives us $\mathfrak{C}\left({\lambda^{-1}}_I, [A,C]\right){\phi^{-1}}_{II,C} = {\phi^{-1}}_{I,C}\mathfrak{C}\left({\lambda^{-1}}_I A, C\right)$. Using this we may finally state

$$ {\phi^{-1}}_{II,C}\left(gf\lambda_A\,(\lambda_I A)\right){\lambda^{-1}}_I\lambda_I = {\phi^{-1}}_{I,C}\left(gf\lambda_A\right)\lambda_I = [gf]\lambda_I, $$

thereby completing the proof of (1).

When considering (2), we first note that we expect both the arrows, via adjunction, to have the form $(([C,D]\otimes[B,C])\otimes[A,B])\otimes A\to D$. Of course, we know that we cannot use ev – essentially the content of the adjoint to internal composition – when the domain is in such a form. Given this, we expect the adjoints of the arrows in (2) to change the domain $(([C,D]\otimes[B,C])\otimes[A,B])\otimes A\to[C,D]\otimes([B,C]\otimes([A,B]\otimes A))$, via some isomorphism.

With this context established, we claim that the following pairs are adjoint to one another.

$$\text{adj}\quad \begin{cases} \circ_{A,B,D}\left(\circ_{B,C,D}\otimes \mathrm{id}_{[A,B]}\right) \\ \circ_{A,C,D}\left(\mathrm{id}_{[C,D]}\otimes\circ_{A,B,C}\right)\alpha_{[C,D],[B,C],[A,B]} \end{cases} \longleftrightarrow \begin{cases} \kappa\alpha_{[C,D],[B,C],[A,B]\otimes A}\alpha_{[C,D]\otimes[B,C],[A,B],A} \\ \kappa(\mathrm{id}_{[C,D]}\otimes\alpha_{[B,C],[A,B],[A]})\alpha_{[C,D],[B,C]\otimes[A,B],A}(\alpha_{[C,D],[B,C],[A,B]}\otimes\mathrm{id}_A) \end{cases}$$

where $\kappa=\mathrm{ev}_{C,D}(\mathrm{id}_{[C,D]}\otimes\mathrm{ev}_{B,C}(\mathrm{id}_{[B,C]}\otimes\mathrm{ev}_{A,B}))$. Were this to be the case, it would immediately follow, by the pentagonal identity (def. (III) 2.1.1), that the right pair were equal and consequently that the left pair were equal.

Thus, in order to prove (2) we must simply prove that we have the isomorphism proposed above. The process is fairly mechanical, and revolves around the naturality of α and ϕ, and makes use of prop. (III) 2.3.6 (4) once in each case. We will show only the first case, and leave the second to the capable hands of the reader.

$\circ_{A,B,D}\left(\circ_{B,C,D}[B,A]\right)$

(defn.)

$$=\phi^{-1}{}_{[B,D][A,B],D}\left[\mathrm{ev}_{B,D}\left([B,D]\,\mathrm{ev}_{A,B}\right)\alpha_{[B,D],[A,B],A}\right]\left(\circ_{B,C,D}[A,B]\right)$$

(naturality of ϕ^{-1})

$$=\phi^{-1}{}_{([C,D][B,C])[A,B],D}\left[\mathrm{ev}_{B,D}\left([B,D]\,\mathrm{ev}_{A,B}\right)\alpha_{[B,D],[A,B],A}\left(\left(\circ_{B,C,D}[A,B]\right)A\right)\right]$$

(naturality of α, we omit subscripts on ϕ^{-1})

$$=\phi^{-1}\left[\mathrm{ev}_{B,D}\left([B,D]\,\mathrm{ev}_{A,B}\right)\left(\circ_{B,C,D}\left([A,B]A\right)\right)\alpha_{[C,D][B,C],[A,B],A}\right]$$

(rearrange composite)

$$=\phi^{-1}\left[\mathrm{ev}_{B,D}\left(\circ_{B,C,D}B\right)\left(([C,D][B,C])\,\mathrm{ev}_{A,B}\right)\alpha_{[C,D][D,C],[A,B],A}\right]$$

(prop. (III) 2.3.6 (4))

$$=\phi^{-1}\left[\mathrm{ev}_{[C,D]}\left([C,D]\,\mathrm{ev}_{B,C}\right)\alpha_{[C,D],[B,C],B}\left(([C,D][B,C])\,\mathrm{ev}_{A,B}\right)\alpha_{[C,D][D,C],[A,B],A}\right]$$

(naturality of α)

$$=\phi^{-1}\Big[\underbrace{\mathrm{ev}_{[C,D]}\left([C,D]\,\mathrm{ev}_{B,C}\right)\left([C,D]\left([B,C]\,\mathrm{ev}_{A,B}\right)\right)}_{\kappa}\alpha_{[C,D],[B,C],[A,B]A}\alpha_{[C,D][D,C],[A,B],A}\Big]$$

The proof of (3) is a similar mechanical exercise, except that we must make use of prop. (III) 2.1.3 (2) or the triangular identity of def. (III) 2.1.1 in order to conclude it.

$$\circ_{A,B,B}\left([\mathrm{id}_B][A,B]\right) = {\phi^{-1}}_{[B,B][A,B],B}\left[\mathrm{ev}_{B,B}\left([B,B]\,\mathrm{ev}_{A,B}\right)\alpha_{[B,B],[A,B],A}\right]\left([\mathrm{id}_B][A,B]\right)$$

(naturality of ϕ^{-1})

$$= {\phi^{-1}}_{I[A,B],B}\left[\mathrm{ev}_{B,B}\left([B,B]\,\mathrm{ev}_{A,B}\right)\alpha_{[B,B],[A,B],A}\left(\left([\mathrm{id}_B][A,B]\right)A\right)\right]$$

(naturality of α)

$$= \phi^{-1}\left[\mathrm{ev}_{B,B}\left([B,B]\,\mathrm{ev}_{A,B}\right)\left([\mathrm{id}_B]([A,B]A)\right)\alpha_{I,[A,B],A}\right]$$

(prop. (III) 2.3.6 (2))

$$= \phi^{-1}\left[\lambda_B\Big(I\big(\underbrace{\mathrm{ev}_{A,B}\left([A,B]A\right)}\big)\Big)\alpha_{I,[A,B],A}\right]$$

($\mathrm{ev}_{A,B}([A,B]\,\mathrm{id}_A) = \mathrm{ev}_{A,B}$ by prop. (A) 1.0.14 and naturality of ϕ)

$$= \phi^{-1}\left[\lambda_B\left(I\,\mathrm{ev}_{A,B}\right)\alpha_{I,[A,B],A}\right]$$

(naturality of λ)

$$= \phi^{-1}\left[\mathrm{ev}_{A,B}\,\lambda_{[A,B]A}\alpha_{I,[A,B],A}\right]$$

(prop. (III) 2.1.3 (2))

$$= \phi^{-1}\left[\mathrm{ev}_{A,B}(\lambda_{[A,B]}A)\right]$$

(prop. (III) 2.3.6 (1))

$$= \lambda_{[A,B]}$$

Again, we give only one of the two equalities, as the other proof is entirely similar – note that instead of requiring prop. (III) 2.1.3 (2) to tie matters together, the other equality makes use of the triangular identity of def. (III) 2.1.1.

Finally, for (4), let $\tau_{-,A,B} = \mathfrak{C}({\alpha^{-1}}_{-,A,B}, C)$ to find

$$\begin{aligned} h_{[A,[B,C]]} &= \mathfrak{C}(-,[A,[B,C]]) \overset{\text{adj}}{\cong} \mathfrak{C}(-\otimes A,[B,C]) \overset{\text{adj}}{\cong} \mathfrak{C}((-\otimes A)\otimes B, C) \\ &\overset{\tau}{\cong} \mathfrak{C}(-\otimes(A\otimes B),C) \overset{\text{adj}}{\cong} \mathfrak{C}(-,[A\otimes B,C]) = h_{[A\otimes B,C]} \end{aligned}$$

Thus, by Yoneda, $[A\otimes B,C] \cong [A,[B,C]]$. ■

In summary then, a right-closed monoidal category affords a rich internal structure. There are internal morphism objects with a well-defined, unital and associative composition law, and the internal morphism objects support the 'same' adjunction formula as do the external ones. Indeed, it would appear that we could reformulate many statements in the general theory to those about *internal* morphisms in some right-closed monoidal category without any loss of generality.

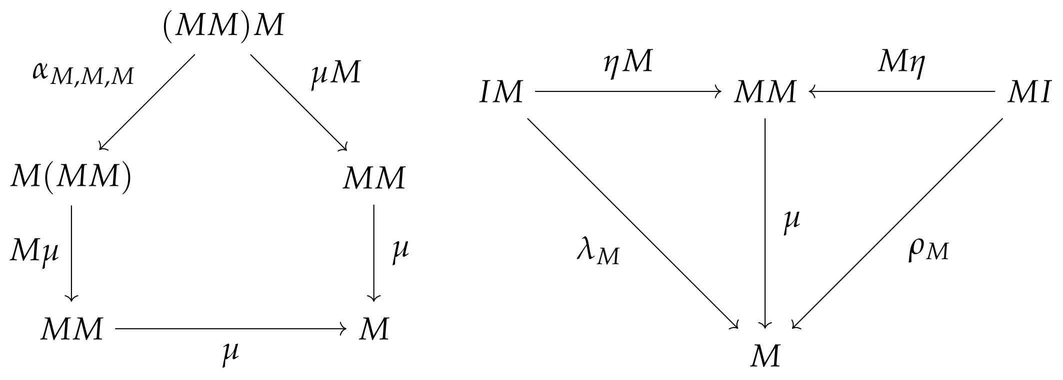
2.4. Monoids in monoidal categories

Now that we are satisfied with the existence of morphisms between monoidal categories, we see that we may form the category of (small) monoidal categories. The reader is encouraged to convince himself that the monoidal nature of the objects and morphisms in this category does not, in any way, prohibit the category from supporting all finite products (in the same manner as does Cat). Thus, among other possible structures, we may endow that category of monoidal categories with a cartesian monoidal structure.

As was pointed out earlier, we think of morphisms from the identity object of a monoidal category to a given object as generalised elements of that object. Ergo, a 'generalised object' of a monoidal category ought to correspond to monoidal functors from the terminal category **1** to the monoidal category in question, as morphisms within the cartesian monoidal category of monoidal categories.

Recall that **1** is the category with only one object and only the identity morphism, and is endowed with a monoidal structure in the obvious and trivial manner. Then, to give a functor from **1** to $\mathfrak{C}$ is to give an object of $\mathfrak{C}$, say $F\star = M$. We are then afforded morphisms $\phi_{\star,\star} : MM \to M$ and $\varepsilon : I \to M$. These are suspiciously reminiscent of multiplication and identity operations in an algebraic monoid. Indeed, should we examine the diagrams in def. (III) 2.1.9 carefully, we find we are able to make the following general definition.

Def. (III) 2.4.1. A monoid in a monoidal category $\mathfrak{C}$ is given by an object $M \in \operatorname{Obj}\mathfrak{C}$ equipped with morphisms $\mu : MM \to M$ and $\eta : I \to M$, known as the multiplication and unit respectively, such that the following diagrams commute.

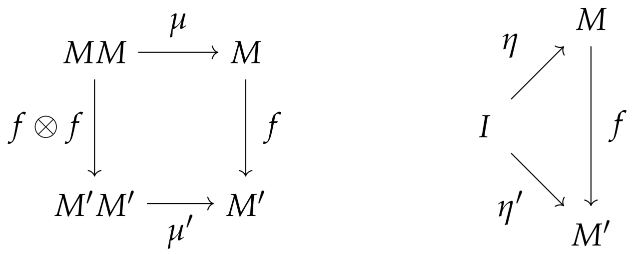


Def. (III) 2.4.2. If the monoidal category has a symmetry β then the monoid is said to be commutative if $\mu\beta = \mu$.

In the above, the left diagram expresses the associativity of the monoidal multiplication while the right expresses the right and left identity laws. It goes without saying that a monoid in the cartesian monoidal category Set is just an algebraic monoid in the usual sense.

If monoids are monoidal functors from the terminal category, then we may guess that monoid morphisms, whatever those may be, are monoidal natural transforms between such functors. Indeed, generalising the resulting diagrams and requirements we define monoid morphisms as follows.

Def. (III) 2.4.3. If (M, μ, η) and (M', μ', η') are monoids in the same monoidal category then a morphism $f : M \to M'$ is a morphism of monoids if the following diagrams commute.



Here, the left diagram makes explicit that multiplication and the morphism should commute, while the right diagram enforces that the morphism take the identity to the identity. As such, in Set, this is simply the definition of an algebraic monoid homomorphism. As a matter of course, one is led to consider the category of monoids on a monoidal category, Mon $\mathfrak{C}$. Given our inspiration for defining monoids, it should come as no surprise that

Prop. (III) 2.4.4. *For any monoidal category $\mathfrak{C}$,* $\mathrm{Mon}\,\mathfrak{C} \cong [\mathbf{1}, \mathfrak{C}]$.

Proof. Consider the maps $\alpha : \mathrm{Mon}\,\mathfrak{C} \to [\mathbf{1}, \mathfrak{C}]$ and $\beta : [\mathbf{1}, \mathfrak{C}] \to \mathrm{Mon}\,\mathfrak{C}$ as defined by $\alpha(M, \mu, \eta) = ([M], [\mu], \eta)$ and $\beta(F, \phi, \varepsilon) = (F\star, \phi_{\star,\star}, \varepsilon)$ on objects, and $\alpha(f) = [f]$ and $\beta(\tau) = \tau_\star$ on arrows. Here we have used the functor $[M]$ where $[M](\star) = M$ and $[M](f) = \mathrm{id}_M$, and the natural transformation $[f]$ which is the constant f natural transformation. It is simple to verify that these maps are indeed functors and inverse to one another, and that they have domain and codomain as stated. ■

Prop. (III) 2.4.5. *Let $(\mathfrak{C}, I_\mathfrak{C})$ and $(\mathfrak{D}, I_\mathfrak{D})$ be monoidal categories, and $F : \mathfrak{C} \to \mathfrak{D}$ be a lax monoidal functor between them. The image of a monoid (M, μ_M, η_M) in $\mathfrak{C}$ has an induced monoid structure in $\mathfrak{D}$. Moreover, such a functor takes monoid morphisms to monoid morphisms in this sense.*

Proof. We shall prove that $(FM, \mu_{FM}, \eta_{FM})$ is a monoid in $\mathfrak{D}$, where we define the arrows $\eta_{FM} = F\eta_M\varepsilon : I_{\mathfrak{D}} \to FM$ and $\mu_{FM} = F\mu_M\phi_{M,M} : FM \otimes FM \to FM$.

We begin with the unitality diagrams. Consider that we wish to demonstrate that $\lambda_{\mathfrak{D}} = \mu_{FM}(\eta_{FM} \otimes \mathrm{id}_{FM})$. To do so, we make use of the diagram for $\lambda_{\mathfrak{D}}$ in def. (III) 2.1.9 which tells us that $\lambda_{\mathfrak{D}} = F\lambda_{\mathfrak{C}}\phi_{I_{\mathfrak{C}},M}(\varepsilon \otimes \mathrm{id}_{FM})$. However, as M was a monoid in $\mathfrak{C}$, we must have that $F\lambda_{\mathfrak{C}} = F\mu_M F(\eta_M \otimes \mathrm{id}_M)$. Now, by naturality of ϕ we have $F(\eta_M \otimes \mathrm{id}_M)\phi_{I_{\mathfrak{C}},M} = \phi_{M,M}(F\eta_M \otimes \mathrm{id}_{FM})$ and thus

$$\begin{aligned}
\lambda_{\mathfrak{D}} &= F\lambda_{\mathfrak{C}}\phi_{I_{\mathfrak{C}},M}(\varepsilon \otimes \mathrm{id}_{FM}) \\
&= F\mu_M F(\eta_M \otimes \mathrm{id}_M)\phi_{I_{\mathfrak{C}},M}(\varepsilon \otimes \mathrm{id}_{FM}) && \text{(monoid in } \mathfrak{C}) \\
&= F\mu_M\phi_{M,M}(F\eta_M \otimes \mathrm{id}_{FM})(\varepsilon \otimes \mathrm{id}_{FM}) && \text{(naturality of } \phi) \\
&= \mu_{FM}\eta_{FM} && \text{(definition)}
\end{aligned}$$

The proof follows, *mutatis mutandis*, for $\rho_{\mathfrak{D}}$.

To see that the associativity holds in $\mathfrak{D}$, we must turn to the associativity diagram for ϕ in def. (III) 2.1.9. Should we paste the image of the associativity diagram for M in $\mathfrak{C}$ to the bottom of that diagram, we find a large commuting diagram allowing two distinct avenues of traversal. In the first case we find

$$\begin{aligned}
&F\mu_M F(\mu_M \otimes \mathrm{id}_M)\phi_{M\otimes M,M}(\phi_{M,M} \otimes \mathrm{id}_{FM}) \\
=&(F\mu_M\phi_{M,M})(F\mu_M \otimes \mathrm{id}_{FM})(\phi_{M,M} \otimes \mathrm{id}_{FM}) && \text{(naturality of } \phi) \\
=&\mu_{FM}(\mu_{FM} \otimes \mathrm{id}_{FM}),
\end{aligned}$$

whereas the second gives

$$\begin{aligned}
&F\mu_M F(\mathrm{id}_M \otimes \mu_M)\phi_{M,M\otimes M}(\mathrm{id}_{FM} \otimes \phi_{M,M})\alpha_{\mathfrak{D}} \\
=&(F\mu_M\phi_{M,M})(F\,\mathrm{id}_M \otimes F\mu_M)(\mathrm{id}_{FM} \otimes \phi_{M,M})\alpha_{\mathfrak{D}} && \text{(naturality of } \phi) \\
=&\mu_{FM}(\mathrm{id}_{FM} \otimes \mu_{FM})\alpha_{\mathfrak{D}}.
\end{aligned}$$

As the large diagram commutes, these two are equal and therefore $(FM, \mu_{FM}, \eta_{FM})$ is a monoid in $\mathfrak{D}$.

Finally, if $f : M \to N$ is a morphism of monoids (M, μ_M, η_M) and (N, μ_N, η_N) in $\mathfrak{C}$, then we seek to show that Ff is a morphism of monoids in $\mathfrak{D}$. Consider that $Ff\eta_{FM} = FfF\eta_M\varepsilon = F(f\eta_M)\varepsilon = F\eta_N\varepsilon = \eta_{FN}$ and that

$$\begin{aligned}
Ff\mu_{FM} &= FfF\mu_M\phi_{M,M} = F(f\mu_M)\phi_{M,M} \\
&= F(\mu_N(f \otimes f))\phi_{M,M} && \text{(monoid morphism in } \mathfrak{C}) \\
&= F\mu_N F(f \otimes f)\phi_{M,M} = F\mu_N\phi_{N,N}(Ff \otimes Ff) && \text{(naturality of } \phi) \\
&= \mu_{FN}(Ff \otimes Ff),
\end{aligned}$$

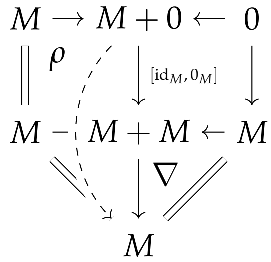
thereby completing the proof. ■

Cor. (III) 2.4.6. *A lax monoidal functor* $F : \mathfrak{C} \to \mathfrak{D}$ *induces a functor* $\mathrm{Mon}\,\mathfrak{C} \to \mathrm{Mon}\,\mathfrak{D}$.

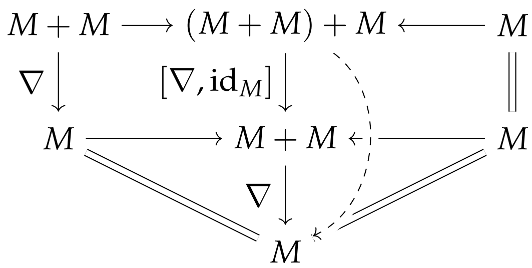
Prop. (III) 2.4.7. *In a cocartesian monoidal category, every object admits a unique commutative monoid structure, and morphisms between objects are morphisms of the induced monoids.*

Proof. Let $\mathfrak{C}$ be have all finite coproducts, fix $M \in \operatorname{Obj} \mathfrak{C}$ and consider $M + M$ in the monoidal category $(\mathfrak{C}, +, 0)$. We obviously have morphisms $\nabla : M + M \to M$ (known as the codiagonal) and $0_M : 0 \to M$ and so only need to show that the diagrams in def. (III) 2.4.1 commute.

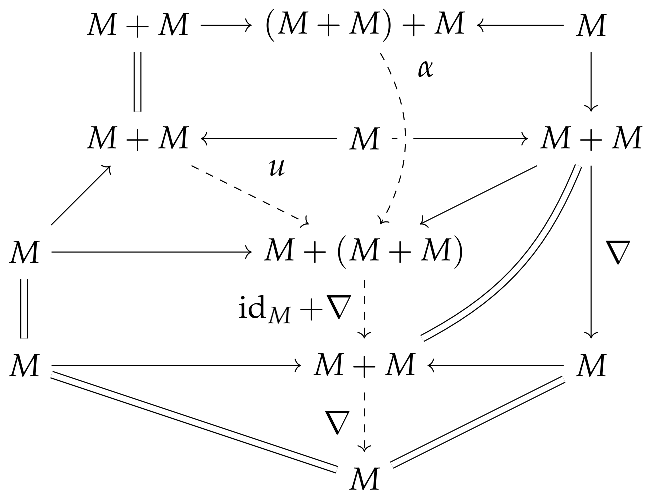
We address first the comparatively short matter of unitality. As is likely evident, that the requisite diagrams commute is due to a simple universal property argument. In particular, to see that $\nabla[\mathrm{id}_M, 0_M] = \rho_M$ consider the following commutative diagram.



The above argument applies, *mutatis mutandis*, to λ. The matter of associativity is more nuanced, but is still simply a universal property argument. In particular, we wish to show that $\nabla[\mathrm{id}_M, \nabla]\alpha = [\nabla, \mathrm{id}_M]\nabla$. Drawing out the appropriate diagram for the right-hand side, we find that it is characterised by the following universal property in the commutative diagram below.



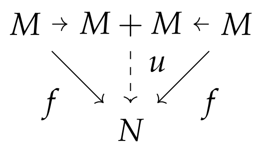
Thus we must verify that $f = \nabla[\mathrm{id}_M, \nabla]\alpha$ has $f\iota_M = \mathrm{id}_M$ and $f\iota_{M+M} = \nabla$ to show the required equality. To see that it does, we draw the diagram giving α, connected appropriately to those for $[\mathrm{id}_M, \nabla]$ and ∇, below. The result follows by noting that $f\iota_M = \nabla\iota_M = \mathrm{id}_M$ (right edge) and $\nabla[\mathrm{id}_M, \nabla]\alpha\iota_{M+M} = \nabla[\mathrm{id}_M, \nabla]u = \nabla\,\mathrm{id}_M$.



Then, that the monoid is commutative is trivial by universal properties again as we know that the symmetry $\beta : M + M \to M + M$ satisfies $\beta\iota_M = \iota_M$ and so $\nabla\beta\iota_M = \iota_M$ gives us $\nabla\beta = \nabla$ by universal property.

To see that the monoid is unique, suppose $\mu : M + M \to M$ was another arrow such that $(M, \mu, 0_M)$ formed a monoid (0_M is obviously unique). As such, it must be the case that $\lambda_M = \mu[0_M, \mathrm{id}_M]$. However, $\lambda_M\iota_M = \mathrm{id}_M$ and so $\mu[0_M, \mathrm{id}_M]\iota_M = \mu_M\iota_M = \mathrm{id}_M$. Thus, by the universal property of ∇, $\mu = \nabla$.

Finally, suppose $f : M \to N$ was an arrow in the category. It is a trivial matter to see that $f0_M = 0_N$ and so we must only check that $f\nabla = \nabla[f, f]$. Once more we apply a standard universal property argument.



Using the above commuting diagram, we see that we have two potential candidates for u, viz., $f\nabla$ and $\nabla[f, f]$. By construction, $f\nabla\iota_M = f$ and it is a simple matter to verify that $\nabla[f, f]\iota_M = f$, thus f is a morphism of monoids. ■

Cor. (III) 2.4.8. *If $\mathfrak{C}$ is a cocartesian monoidal category, then $\mathfrak{C} \cong \mathrm{Mon}\,\mathfrak{C}$.*

3. Enriched categories

A recurring theme in the theory of categories is that meaningful results and information can be obtained not by studying the constituents of any given object, but rather by studying the interdependence that the object has with other, related objects. That is, in a category it is the arrows that are in some sense more important than the 'elements' of a given object, should such a notion even exist. Given this, it is curious then that the collection of morphisms between two objects be defined to comprise individual elements – it has, in a way, a privileged position. Moreover, it seems at odds with the rampant generalisation present elsewhere that we should be forced to deal with categories whose collections of morphisms are confined to form sets (or classes) and not other interesting structures – groups, topological spaces, and even categories themselves! To remedy these shortcomings, we introduce the notion of enriched categories.

3.1. Basic notions

Def. (III) 3.1.1. Let $(\mathfrak{V}, \otimes, I, \alpha, \lambda, \rho)$ be a monoidal category, a $\mathfrak{V}$-enriched category or $\mathfrak{V}$-category, $\mathfrak{C}$ is a collection of objects $\operatorname{Obj}\mathfrak{C}$ such that

1. for each ordered pair of objects $(A, B) \in \operatorname{Obj}\mathfrak{C} \times \operatorname{Obj}\mathfrak{C}$ there is an associated object $\mathfrak{C}(A, B) \in \operatorname{Obj}\mathfrak{V}$, called the morphism object from A to B
2. for each ordered triple $(A, B, C) \in \operatorname{Obj}\mathfrak{C}^{\times 3}$ there is a morphism $\circ_{A,B,C} \in \operatorname{Mor}\mathfrak{V}$ with $\circ_{A,B,C} : \mathfrak{C}(B, C) \otimes \mathfrak{C}(A, B) \to \mathfrak{C}(A, C)$, called the composition morphism
3. for each object $A \in \operatorname{Obj}\mathfrak{C}$ there is a morphism $j_A : I \to \mathfrak{C}(A, A)$, called the identity element

where the following diagrams must commute for all $A, B, C, D \in \operatorname{Obj}\mathfrak{C}$, expressing the associativity of composition and that composition is unital, respectively.

$$\begin{array}{ccc}
(\mathfrak{C}(C,D)\otimes\mathfrak{C}(B,C))\otimes\mathfrak{C}(A,B) & \xrightarrow{\alpha_{\mathfrak{C}(C,D),\mathfrak{C}(B,C),\mathfrak{C}(A,B)}} & \mathfrak{C}(C,D)\otimes(\mathfrak{C}(B,C)\otimes\mathfrak{C}(A,B)) \\
\Big\downarrow{\scriptstyle \circ_{B,C,D}\otimes\mathrm{id}_{\mathfrak{C}(A,B)}} & & \Big\downarrow{\scriptstyle \mathrm{id}_{\mathfrak{C}(C,D)}\otimes\circ_{A,B,C}} \\
\mathfrak{C}(B,D)\otimes\mathfrak{C}(A,B)\xrightarrow[\circ_{A,B,D}]{} & \mathfrak{C}(A,D) & \xleftarrow[\circ_{A,C,D}]{}\mathfrak{C}(C,D)\otimes\mathfrak{C}(A,C)
\end{array}$$

$$\begin{array}{ccccc}
\mathfrak{C}(B,B)\otimes\mathfrak{C}(A,B) & \xrightarrow{\circ_{A,B,B}} & \mathfrak{C}(A,B) & \xleftarrow{\circ_{A,A,B}} & \mathfrak{C}(A,B)\otimes\mathfrak{C}(A,A) \\
\Big\uparrow{\scriptstyle j_B\otimes\mathrm{id}_{\mathfrak{C}(A,B)}} & \nearrow{\scriptstyle \lambda_{\mathfrak{C}(A,B)}} & & \nwarrow{\scriptstyle \rho_{\mathfrak{C}(A,B)}} & \Big\uparrow{\scriptstyle \mathrm{id}_{\mathfrak{C}(A,B)}\otimes j_A} \\
I\otimes\mathfrak{C}(A,B) & & & & \mathfrak{C}(A,B)\otimes I
\end{array}$$

Perhaps the simplest, non-trivial example of a $\mathfrak{V}$-category is that of an enriched singleton set.

Example (III) 3.1.2

> Let $\text{Obj}\,\mathfrak{C} = \{\star\}$, and consider $\mathfrak{C}$ as a $\mathfrak{V}$-category. As such, $\mathfrak{C}(\star, \star) = M \in \text{Obj}\,V$ is a single distinguished object with $j : I \to M$ and $\circ : M \otimes M \to M$, where the appropriate diagrams commute. Careful inspection reveals these diagrams to be precisely those present in def. (III) 2.4.1 and so an enriched singleton is precisely a monoid object in the underlying monoidal category.

To demonstrate that enriched categories are indeed a generalisation of standard categories, we note the following two cases of interest.

Example (III) 3.1.3

> If we take $\mathfrak{V}$ to be the cartesian monoidal category Set, then a $\mathfrak{V}$-category can be seen simply as a locally small category. If we are more daring and set $\mathfrak{V}$ to be the cartesian monoidal category Cat of small categories, then we recover a 2-category.

Example (III) 3.1.4

> Finally, with an eye to closed categories, we note that if $\mathfrak{V}$ is a right closed monoidal category then it is canonically enriched over itself, with $\mathfrak{V}(A, B)$ defined to be $[A, B]$, $j_A = [\text{id}_A]$ and $\circ$ as before. That the relevant diagrams commute has already been shown in prop. (III) 2.3.7. Thus, Set, Ab, Cat are all enriched over themselves.

Def. (III) 3.1.5. A functor between $\mathfrak{V}$-categories $F : \mathfrak{C} \to \mathfrak{D}$, a $\mathfrak{V}$-functor, is given by a set map $F : \text{Obj}\,\mathfrak{C} \to \text{Obj}\,\mathfrak{D}$ together a morphism $F_{A,B} : \mathfrak{C}(A, B) \to \mathfrak{D}(FA, FB)$ in $\mathfrak{V}$ for each $A, B \in \text{Obj}\,\mathfrak{C}$, such that the following diagrams commute.

$$\begin{array}{ccc} \mathfrak{C}(B,C) \otimes \mathfrak{C}(A,B) & \xrightarrow{\circ_{A,B,C}} & \mathfrak{C}(A,C) \\ {\scriptstyle F_{B,C} \otimes F_{A,B}} \downarrow & & \downarrow {\scriptstyle F_{A,C}} \\ \mathfrak{D}(FB,FC) \otimes \mathfrak{D}(FA,FB) & \xrightarrow[\circ_{FA,FB,FC}]{} & \mathfrak{D}(FA,FC) \end{array}$$

$$\begin{array}{ccc} & I & \\ {\scriptstyle j_A} \swarrow & & \searrow {\scriptstyle j_{FA}} \\ \mathfrak{C}(A,A) & \xrightarrow{F_{A,A}} & \mathfrak{D}(FA,FA) \end{array}$$

The above diagrams simply serve to indicate that a $\mathfrak{V}$-category functor must respect identity and composition, as we would have it in the standard case. Of course, setting $\mathfrak{V} = \text{Set}$, we recover the standard definition of a functor.

Now that we have functors between $\mathfrak{V}$-categories, we may be tempted to contrive the definition of natural transformations between such categories. In the case of $\mathfrak{V}$=Set, we understand a natural transform $\eta : F \to G$ between functors $F, G : \mathfrak{C} \rightrightarrows \mathfrak{D}$ to be a collection of arrows $\eta_A : FA \to GA$ for each object $A \in \operatorname{Obj}\mathfrak{C}$. However, in the enriched context we cannot directly speak of individual arrows. As such, we employ the 'trick' of instead giving an arrow $\eta_A : I \to \mathfrak{D}(FA, GA)$ and specifying its properties so that it serves as though it were 'choosing' the correct morphism, were a map.

Def. (III) 3.1.6. Given two $\mathfrak{V}$-category functors $F, G : \mathfrak{C} \rightrightarrows \mathfrak{D}$, a $\mathfrak{V}$-natural transform $\eta : F \to G$ is given by a family of arrows $\eta_A : I \to \mathfrak{D}(FA, GA)$ indexed by $\operatorname{Obj}\mathfrak{C}$ such the the following diagram commutes, for all $A, B \in \operatorname{Obj}\mathfrak{C}$.

$$\begin{array}{ccccc} I\otimes\mathfrak{C}(A,B) & \xleftarrow{\lambda^{-1}_{\mathfrak{C}(A,B)}} & \mathfrak{C}(A,B) & \xrightarrow{\rho^{-1}_{\mathfrak{C}(A,B)}} & \mathfrak{C}(A,B)\otimes I \\ {\scriptstyle \eta_B\otimes F_{A,B}}\downarrow & & & & \downarrow{\scriptstyle G_{A,B}\otimes\eta_A} \\ \mathfrak{D}(FB,GB)\otimes\mathfrak{D}(FA,FB) & & & & \mathfrak{D}(GA,GB)\otimes\mathfrak{D}(FA,GA) \\ & \searrow{\scriptstyle \circ_{FA,FB,GB}} & & \swarrow{\scriptstyle \circ_{FA,GA,GB}} & \\ & & \mathfrak{D}(FA,GB) & & \end{array}$$

If $\mathfrak{V} = $ Set, then we recover the standard definition of a natural transform, viz., that it must commute with functorial images of arrows. To see this, recall that in Set, $I = \{\star\}$, and so to give η_A is to give a single arrow $FA \to GA$. Then, if we trace out the commutative diagram, beginning with an $f \in \mathfrak{C}(A, B)$, we find the requirement $\eta_B Ff = Gf\eta_A$ – precisely the familiar naturality square.

There is much that can be said for the theory of enriched categories – for example, we may attempt to recast all of the results of the standard theory in the enriched setting. By and large, this has been done (enriched adjunctions, limits, Yoneda, *etc.*) and the results have had a profound influence on the direction of the theory and formulation of the "higher category" theory. For lack of time and direct applicability to later sections, the author regrets that such directions have not been included in this work.

In what follows, we will examine some select examples of enrichment which will play a central role in later chapters.

3.2. Semi-additive categories

Now that the defining facets of enriched categories have been made clear, we turn our attention to a particular case of enrichment, viz., categories enriched over $(\textsc{Set}_\bullet, \wedge)$ where $\wedge$ is the smash product of spaces[2].

Should we work carefully through the diagrams in def. (III) 3.1.1, we see that a category enriched over $\textsc{Set}_\bullet$ has a distinguished element in each morphism object, and that composition takes distinguished elements to distinguished elements. Due to our algebraic inclinations, we say that such distinguished morphisms are called zero morphisms, and abstracting, we reach the following definition.

Def. (III) 3.2.1. A category has zero morphisms if $(\forall A, B \in \operatorname{Obj}\mathfrak{C})\exists 0_{AB} \in \mathfrak{C}(A, B)$ such that the following diagram commutes for all $A, B, C \in \operatorname{Obj}\mathfrak{C}$ and all $f : A \to B$ and $g : B \to C$.

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ 0_{AB}\big\downarrow & \searrow^{0_{AC}} & \big\downarrow 0_{BC} \\ B & \xrightarrow[g]{} & C \end{array}$$

That is, there is a system of morphisms which are biconstant in a compatible way.

Prop. (III) 3.2.2. *Zero morphism systems are unique if they exist.*

Proof. Let 0 and $0'$ be two systems of zero morphisms over a category. Consider that for all objects A, B, C we must have $0_{A,C} = 0'_{B,C}0_{A,B} = 0'_{A,C}$. ■

It may be observed that every category with zero morphisms can be seen as enriched over $(\textsc{Set}_\bullet, \wedge)$, including specifically $\textsc{Set}_\bullet$. Although such an enrichment is a structural property, we may reach it through an entirely different avenue.

Def. (III) 3.2.3. The zero object of a category, should it exist, is an object that is both initial and terminal.

Prop. (III) 3.2.4. *A category with a zero object has zero morphisms.*

Proof. The proof is trivial as every arrow factors through the zero object, and universal properties necessitate the rest. ■

Thus, the presence of a particular *object* in a category can determine a *structural* property. Moreover, we have a partial converse in the presence of specific morphisms.

Prop. (III) 3.2.5. *In a category with zero morphisms, the following are equivalent:*

1. *There is a zero object*
2. *There is a terminal object*
3. *There is an initial object.*

[2] Recall that smash product is defined as $(A, a_0) \wedge (B, b_0) = A \times B/\sim$ where $(a, b_0) \sim (a_0, b)$.

Proof. It is clear that (1) implies (2) and (3). We show only that (2) implies (1), the other implication follows by dualisation.

Let 1 be the terminal object. That for every object C in the category there exists a morphism $1 \to C$ is clear by the existence of zero morphisms. We need only show that $0_{1,C}$ is unique. To that end, let $f : 1 \to C$ be an arrow in the category. Recall that $\mathfrak{C}(1,1) = \{\mathrm{id}_1\}$ and so $f = f\,\mathrm{id}_1 = f0_{1,1} = 0_{1,C}$, thus 1 is initial. ■

In order to drive home the point that there is no full converse to prop. (III) 3.2.4,

Non-example (III) 3.2.6

> Consider a ring as a monoid under multiplication and view it as a one-object category. This category has a zero morphism, but no initial or terminal objects.

Remark (III) 3.2.7. In homage to its enriched heritage, we say that a category enriched over $\mathrm{SET}_\bullet$ which has a zero object is a *pointed category.*

A remarkable property of categories with zero morphisms (and so of pointed categories) is the existence of a very special morphism from the coproduct of a collection of objects, to the product of that same collection, when both exist. In order to enable effective discussion of this, we make the following small definition.

Def. (III) 3.2.8. In a category with zero morphisms, for any pair of objects A, B, define $\delta_{A,A} = \mathrm{id}_A$ and $\delta_{A,B} = 0_{A,B}$ when $A \neq B$. When a collection of objects $(C_i)_{i\in I}$ is considered, we write $\delta_{j,k}$ for δ_{C_j,C_k}.

Prop. (III) 3.2.9. *In a category with zero morphisms, if the collection of objects $(C_i)_{i\in I}$ has both a product and a coproduct, then there exists a unique morphism $\alpha : \coprod C_i \to \prod C_i$ such that $\pi_k \alpha \iota_i = \delta_{i,k}$.*

Proof. For each $k \in I$ we have a unique arrow $[(\delta_{ik})_{i\in I}] : \coprod C_i \to C_k$ such that $[(\delta_{ik})_{i\in I}]\iota_j = \delta_{jk}$. With these arrows we define $\alpha = \langle([(\delta_{ij})_{i\in I}])_{j\in I}\rangle : \coprod C_i \to \prod C_i$ to be the unique arrow with projections $\pi_k\alpha = [(\delta_{ik})_{i\in I}]$. Uniqueness follows easily by universal property. ■

The observant reader will here notice

Cor. (III) 3.2.10. *In a category with zero morphisms, if the collection of objects $(C_i)_{i\in I}$ has both a product and a coproduct, then*

$$\left\langle \left([(\delta_{ij})_{i\in I}]\right)_{j\in I} \right\rangle = \left[\left(\langle(\delta_{ij})_{j\in I}\rangle\right)_{i\in I} \right] : \coprod C_i \to \prod C_i$$

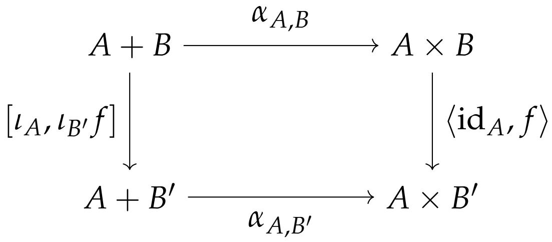
Later we shall see a sense in which this statement is obviously true, but for the time being we allow the further exploration of the properties (desired and inherent) of α to guide us onward.

A first inroad into the properties of α may be that of asking how it 'transforms' as the underlying components of the (co)product change under morphisms. More directly, we may wish to know whether, for binary (co)products, α is natural.

Prop. (III) 3.2.11. *In a category with zero morphisms, if all pairs of objects admit a product and a coproduct, then $\alpha_{A,B} : A + B \to A \times B$ as defined in prop. (III) 3.2.9 is a natural transformation between the bifunctors $\times, + : \mathfrak{C} \times \mathfrak{C} \rightrightarrows \mathfrak{C}$.*

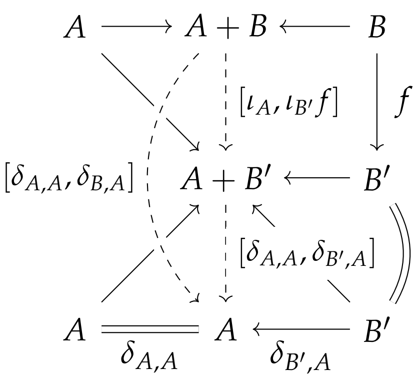
Proof. We prove only naturality in the second argument explicitly here, as naturality in the first is entirely similar. Then, a natural transform that is independently natural in both arguments is binatural and the proof is completed.

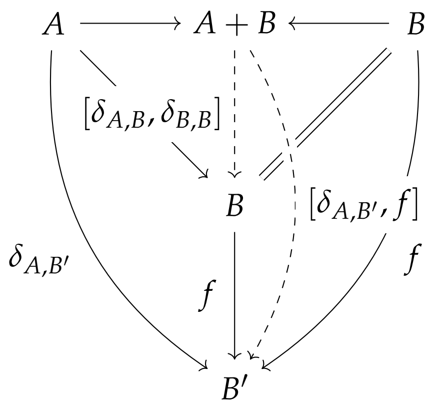
To this end, fix objects A, B and arrow $f : B \to B'$ in the category, we desire that the following diagram commute.



Given that we have $g = [\delta_{A,A}, \delta_{B,A}] : A + B \to A$ and $h = [\delta_{A,B'}, f] : A + B \to B'$ we must have a unique arrow $u : A + B \to A \times B'$ such that $\pi_A u = g$ and $\pi_{B'} u = h$. However, we have two potential candidates, viz., $\langle \mathrm{id}_A, f \rangle \alpha_{A,B}$ and $\alpha_{A,B'}[\iota_A, \iota_{B'} f]$. To ensure that they are both candidates, we must check that they satisfy the above-mentioned identities. This is immediate in all cases, but to elucidate matters, we expand $\pi_A \alpha_{A,B'}[\iota_A, \iota_{B'} f]$, $\pi_A \langle \mathrm{id}_A, f \rangle \alpha_{A,B}$ and $\pi_B \langle \mathrm{id}_A, f \rangle \alpha_{A,B}$.

For the first, by prop. (III) 3.2.9 we have $\pi_A \alpha_{A,B'}[\iota_A, \iota_{B'} f] = [\delta_{A,A}, \delta_{B',A}][\iota_A, \iota_{B'} f]$ but the right-hand side is equal to $[\delta_{A,A}, \delta_{B,A}] = g$ as the diagram below left commutes (the canonical inclusion maps are unlabelled, and note that $\delta_{B'A} f = \delta_{B,A}$).





That $f[\delta_{A,B}, \delta_{B,B}] = [\delta_{A,B'}, f] = h$ follows from the above-right commuting diagram and so gives us $\pi_B \langle \mathrm{id}_A, f \rangle \alpha_{A,B} = \pi_B \langle [\delta_{A,A}, \delta_{B,A}], f[\delta_{A,B}, \delta_{B,B}] \rangle = f[\delta_{A,B}, \delta_{B,B}] = h$. It is a comparatively simple matter to see that

$$\pi_A \langle \mathrm{id}_A, f \rangle \alpha_{A,B} = \pi_A \langle [\delta_{A,A}, \delta_{B,A}], f[\delta_{A,B}, \delta_{B,B}] \rangle = [\delta_{A,A}, \delta_{B,A}] = g$$

That the final identity, $\pi_B \alpha_{A,B'}[\iota_A, \iota_{B'} f] = h$, holds can be seen from a universal property argument entirely similar to the one given in the above-left diagram. ∎

We can, of course, require even more of our special morphism.

Def. (III) 3.2.12. In a category with zero morphisms, if the unique arrow given in prop. (III) 3.2.9, $\alpha : \coprod C_i \to \prod C_i$, is an ismorphism then we say that $(C_i)_{i\in I}$ admit a biproduct, and write $\bigoplus C_i$ for its product and coproduct.

Remark (III) 3.2.13. There is a subtlety here which bears expanding. If α is indeed an isomorphism, then we understand $\prod C_i \cong \coprod C_i$, but only $\prod C_i$ is equipped with projections π_i and only $\coprod C_i$ is equipped with inclusions ι_i. To this end, when we write $(\bigoplus C_i, \pi_i, \iota_i)$ we understand there to be some slight of hand, as it is not the case that the domain of π_i is *equal* to the codomain of ι_i – somewhere, we must account for α. It is a simple matter to see that if we define $\pi_i : \bigoplus C_j \to C_i$ as π_i of $\prod C_j$ and $\iota_i : C_i \to \bigoplus C_j$ as $\alpha\iota_i$ where the ι_i are from $\coprod C_i$ then we have $\pi_i\iota_j = \delta_{ij}$. Importantly, if we were to define matters the other way around, the equality would still hold. Thus, in some sense, which of $\prod C_i$ and $\coprod C_i$ we set to be equal to $\bigoplus C_i$ does not change the relationship that the inclusions and projections of the biproduct have.

Def. (III) 3.2.14. In a category with zero morphisms, if all finite collections of objects admit biproducts then the category is called semi-additive.

Example (III) 3.2.15

It is easy to see that finite products and coproducts of commutative groups (and monoids) coincide, and that Ab and CMon both have zero morphisms and have all finite biproducts.

Remark (III) 3.2.16. The reason here that we choose to require the existence of only finite (as opposed to arbitrary) biproducts is so that we restrict ourselves to a reasonable generalisation of 'algebraic' categories (examples above). In particular, semi-additive categories will later lead to additive categories and later still will inspire abelian categories whose very design, in so far as we are concerned, is inspired by the desire to support homological theories in a unified manner.

In particular then, all semi-additive categories are pointed. We know that we may also restate the above as the category admitting binary biproducts and having an initial (equivalently terminal) object. With these two definitions, we are ready to prove yet another interesting case of the presence of particular objects providing a global structure.

Prop. (III) 3.2.17. *Every semi-additive category is canonically enriched over* CMon, *the category of commutative algebraic monoids with the canonical cartesian monoidal structure.*

Proof (Sketch). In order to prove this, we need to demonstrate that every set of morphisms admits an algebraic commutative monoid structure which is preserved by composition.

To begin then, recall the results of prop. (III) 2.4.7 and its dual statement. Then, let $\mathfrak{C}$ be semi-additive and fix two objects in the category, A and B, and consider $\mathfrak{C}(A, B)$. We wish to define addition, so take $f, g \in \mathfrak{C}(A, B)$ and define $f + g \in \mathfrak{C}(A, B)$ to be the composite

$$A \xrightarrow{\Delta} A \oplus A \xrightarrow{f \oplus g} B \oplus B \xrightarrow{\nabla} B$$

The commutativity of addition immediately follows from the commutativity of either the comonoid structure on A or monoid structure on B. The associativity is a trivial consequence of the fact that the monoidal structures in question have an associator, coupled with a standard universal property argument. That addition respects $0_{A,B}$ is again the result of a universal property argument involving λ and ρ.

Finally, we must show that composition is a monoid homomorphism from the biproduct. We already know that it preserves $0_{A,B}$ by definition, so we must only check that $(f+g)h = fh + fg$ and $h(f+g) = hf + hg$, but again these follow from universal property arguments based on ∇ and Δ respectively. The curious reader is encouraged to explore the diagrams, but we will not belabour the proof here. ■

Remark (III) 3.2.18. There is one lamentable aspect of this otherwise glorious result. Due to the cartesian monoidal structure of CMon, in a category simply enriched over CMon (and not semi-additive) composition is not required to be bilinear (over $\mathbb{N}$) in the sense that we do not automatically inherit zero morphisms from such an enrichment. This will not be a problem with Ab later, but adds an extra factor to consider here.

Now that we have established addition of morphisms canonically, we will attempt to add a variety of morphisms – especially those relating to biproducts.

Prop. (III) 3.2.19. *In a semi-additive category, if the finite collection C_i admits a biproduct $C = \bigoplus C_i$, then $\sum \iota_i \pi_i = \mathrm{id}_C$.*

Proof. The proof is a straightforward universal property argument. ■

Before we proceed with a generalisation of the above argument, we consider here the nature of the commutative monoids themselves.

Prop. (III) 3.2.20. *In the canonical enrichment of a semi-additive category over* CMon, *the morphism monoids are cancellative.*

Proof. In such a category $\mathfrak{C}$ we wish to show that $a + b = a + c \implies b = c$ for all $a, b, c \in \mathfrak{C}(A, B)$ for any objects $A, B \in \operatorname{Obj} \mathfrak{C}$. To do so, we recall that we defined $a + b = \nabla(a \oplus b)\Delta$ and leverage universal properties to our advantage.

Observe that $\Delta = \iota_0 + \iota_1$ by universal property of biproduct and $\pi_j \iota_i = \delta_{ij}$ by biproduct property. Thus, $\Delta(\pi_0 + \pi_1) = \iota_0\pi_0 + \iota_1\pi_1 + \iota_0\pi_1 + \iota_1\pi_0 = \mathrm{id} + 0$ by the biproduct property and prop. (III) 3.2.19. Ergo by precomposition with $\pi_0 + \pi_1$, that $\nabla(a \oplus b)\Delta = \nabla(a \oplus c)\Delta$ implies $\nabla(a \oplus b) = \nabla(a \oplus c)$. A universal property argument shows that $\nabla(a \oplus c)\iota_1 = c$ and so that $\nabla(a \oplus b) = \nabla(a \oplus c)$ implies $b = c$ by precomposition with ι_1. ■

Now we return to the context of prop. (III) 3.2.19 and show a small, perhaps trivial, but nevertheless consequential generalisation that states that an arrow between biproducts is completely determined by its actions on components.

Prop. (III) 3.2.21. *In a semi-additive category, a morphism $f : \bigoplus_1^n A_i \to \bigoplus_1^m B_i$ is uniquely determined by the nm-many morphisms ${f^i}_j = \pi'_i f \iota_j$, where $(\bigoplus_1^n A_i, \pi', \iota')$ and $(\bigoplus_1^m B_i, \pi, \iota)$ are finite biproducts.*

Proof. This is a consequence of universal properties – specifying the collection $\pi_i f$ determines f uniquely into $\bigoplus B_i$ from its projection onto each B_i. Then, specifying ${f^i}_j = \pi_i f \iota_j$ for fixed i and determines $\pi_i f$ uniquely from $\bigoplus A_i$ from $\pi_i f \iota_j$ on A_j. ■

This result suggests of itself something with which we are very familiar. Indeed, the notation was chosen so as to all but prove the following.

Prop. (III) 3.2.22. *In a semi-additive category, let $A = \bigoplus A_i$, $B = \bigoplus B_i$, and $C = \bigoplus C_i$ be finite biproducts with arrows $f, g : A \rightrightarrows B$ and $h : B \to C$. Then ${(hf)^i}_j = \sum_k {h^i}_k {f^k}_j$ and ${(g+f)^i}_j = {g^i}_j + {f^i}_j$.*

Proof. Let $(A, \pi'', \iota''), (B, \pi', \iota'), (C, \pi, \iota)$ be the finite biproducts in question. To see the first result, consider that

$$\sum_k {h^i}_k {f^k}_j = \sum_k \pi''_i h \iota'_k \pi'_k f \iota_j = \pi''_i h \left(\sum_k \iota'_k \pi'_k \right) f \iota_j = \pi''_i h \, \mathrm{id}_B \, f \iota_j = {(hf)^i}_j$$

where the penultimate equality is due to prop. (III) 3.2.19, and the antepenultimate one is due to prop. (III) 3.2.17. The second result is universal property argument coupled with the distributivity of composition. ■

We have suddenly arrived at something that the reader is reasonably expected to find surprising, should he not have encountered it before. Semi-additive categories lend to their morphisms a calculus of matrices! That is, props. (III) 3.2.21 and (III) 3.2.22 combine to allow us to specify morphisms involving finite biproducts as matrices, where we have extended the notation in the obvious manner as indicated below on the left- and right- most arrows, and where composites correspond to matrix products and parallel sums to matrix sums.

$$A \xrightarrow{\begin{pmatrix} f_1 & \cdots & f_n \end{pmatrix}} \bigoplus^n B_i \xrightarrow{\begin{pmatrix} {g^1}_1 & \cdots & {g^1}_m \\ \vdots & \ddots & \vdots \\ {g^n}_1 & \cdots & {g^n}_m \end{pmatrix}} \bigoplus^m C_i \xrightarrow{\begin{pmatrix} h^1 \\ \vdots \\ h^m \end{pmatrix}} D$$

With this new understanding we return to cor. (III) 3.2.10 and observe that it is simply the statement that specifying the contents of a matrix in row-major or column-major order does not change the matrix *en masse*. Moreover, by omitting a few subscripts, we can recast prop. (III) 3.2.17 in matrix terms to discover that it essentially showed the 'obvious' statement $f + g = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} f & 0 \\ 0 & g \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Further still, prop. (III) 3.2.19 is the general version of the statement that $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

This is, of course, extremely exciting and interesting and the topic would appear to be exploding with questions, the most obvious of which is perhaps: "FinVect$_{\mathbb{F}}$ has biproducts, and FinVect$_{\mathbb{F}}$ is certainly enriched over CMon – do we recover standard matrix linear algebra in this fashion?" and "is there a reasonable generalisation of conjugate transpose?"The answer in both cases is, astoundingly, yes! Regrettably, however, we shall not explore such avenues as they would lead us far astray.

We have seen then, how requiring the existence of all finite biproducts gives us, chiefly, a canonical enrichment over CMon. Conversely, if we begin with an enrichment over CMon, we may define biproducts as follows.

Def. (III) 3.2.23. The biproduct of objects A_0, A_1 is an object B equipped with morphisms $\pi_i : B \to A_i$ and $\iota_i : A_i \to B$ such that $\pi_i \iota_j = \delta_{ij}$ and $\sum \iota_i \pi_i = \mathrm{id}_B$.

This definition is entirely compatible with our earlier definition when the two are applicable, such as in the case of the canonical enrichment of semi-additive categories. Thus, it is no surprise that

Prop. (III) 3.2.24. *In a* CMon*-enriched category with zero morphisms, the existence of the following are equivalent:*

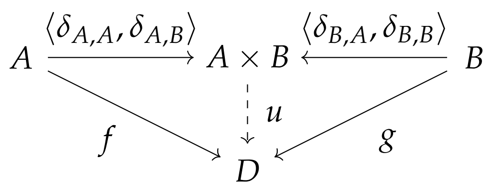
1. *All finite products*
2. *All finite coproducts*
3. *All finite biproducts*

Proof. We begin with (3) implying (1) and (2). The nullary case is obviously true, and we will show only that (3) implies (1), as the the other result is achieved by dualisation. Thus, assume that the biproduct B of the finite collection $(A_i)_{i \in I}$ exists. Given maps $f_i : C \to A_i$ we may form $u = \sum \iota_i f_i$ as an arrow $u : C \to B$ with $\pi_i u = f_i$. To see that u is unique, suppose there was an arrow $v : C \to B$ with $\pi_i v = f_i$, then $v = \mathrm{id}_B\, v = (\sum \iota_i \pi_i) v = (\sum \iota_i f_i) = u$.

Next we show that (1) implies (3), the other result follows by duality. Moreover, our proof will be for the two object case, as this implies the finite case as we have already the nullary case (prop. (III) 3.2.5).

Fix objects A and B in the category, and consider that we have two canonical endomorphisms of $A \times B$, viz., $\langle \delta_{A,A}, \delta_{B,A} \rangle \pi_A$ and $\langle \delta_{B,A}, \delta_{B,B} \rangle \pi_B$. As such, we may form the sum $\sigma = \langle \delta_{A,A}, \delta_{B,A} \rangle \pi_A + \langle \delta_{B,A}, \delta_{B,B} \rangle \pi_B$ and observe that $\pi_A \sigma = \pi_A$ and $\pi_B \sigma = \pi_B$ allowing us to conclude that $\sigma = \mathrm{id}_{A \times B}$ – this will be crucial.

Now suppose that we have an object D with morphisms $f : A \to D$ and $g : B \to D$. We wish to find a unique morphism u such that the following diagram commutes, making $A \times B$ isomorphic to $A + B$



There certainly *exists* a u with the required properties, $u = f\pi_A + g\pi_B$. For uniqueness, suppose there was a $v : A \times B \to D$ and consider that $v\sigma = u\sigma$ by the required properties of v, but $\sigma = \mathrm{id}_{A \times B}$ so $v = u$. ∎

As a last result for this section, not only is the above-define biproduct thus identical, but it even supports the same addition of morphisms.

Prop. (III) 3.2.25. *In a* CMon*-enriched category with biproducts, the sum of the parallel arrows $f, g : A \rightrightarrows B$ admits the identity $f + g = \nabla_B (f \oplus g) \Delta_A$.*

Proof. Observe that $\Delta_A = \iota_0 + \iota_1$ by universal property so that $\nabla_B (f \oplus g) \Delta_A = \nabla_B (f \oplus g)(\iota_0 + \iota_1) = \nabla_B (f \oplus g) \iota_0 + \nabla_B (f \oplus g) \iota_1 = \nabla_B \iota_0 f + \nabla_B \iota_1 g = f + g$. ∎

3.3. Kernels and co.

Continuing in the algebraic vein ushered in by the previous section, we introduce some important generalisations of essentially algebraic notions so as to provide a uniform means to discuss later concepts.

Def. (III) 3.3.1. Given $A \in \operatorname{Obj}\mathfrak{C}$ and $M \subseteq \operatorname{Mono}\mathfrak{C}$ a class of monomorphisms, an M-subobject is an isomorphism class of M-monomorphisms $m : B \to A$. That is, two M-monomorphisms $m : B \to A$ and $m' : B' \to A$ are equivalent iff there exists an isomorphism $k : B \to B'$ such that $m = m'k$. If $M = \operatorname{Mono}\mathfrak{C}$ then we say that $m : B \to A$ is a subobject.

Surprise (III) 3.3.2

In Set, a subobject B of a set A is the class of all injections $m : B' \to A$ such that $|B'| = |B|$, and so not any *subset* in particular. Indeed, it is certainly possible for $B' \cap A = \phi$ in general. Moreover, in Top we see that while subspaces are certainly subobjects, so is the space itself with any finer topology. Thus, subobjects do not necessarily capture the correct notion of containment that we desire when we speak of subsets, subspaces and so on. As such, we will confine future discussion to regular subojects, where $M = \operatorname{RegMono}\mathfrak{C}$.

Def. (III) 3.3.3. Given $A \in \operatorname{Obj}\mathfrak{C}$ and a class of epimorphisms $E \subseteq \operatorname{Epi}\mathfrak{C}$, an E-quotient object is an isomorphism class of epimorphisms $e : A \to B$. That is, two E-epimorphisms $e : A \to B$ and $e' : A \to B'$ are equivalent iff there exists an isomorphism $k : B \to B'$ such that $e = ke'$.

Surprise (III) 3.3.4

In Mon, we find that the inclusion map $\mathbb{N} \overset{\subset}{\to} \mathbb{Z}$ is actually an epimorphism (though *not* a surjection of sets) and so $\mathbb{Z}$ is a quotient object of $\mathbb{N}$. Again then, there is a problem with simply taking all morphisms, and so we restrict attention to regular quotient objects.

With sub- and quotient- objects defined, we are tempted to generalise the standard algebraic examples of such objects.

Def. (III) 3.3.5. In a category with zero morphisms, we define the kernel of a map $f : A \to B$ to be the equaliser of f and 0_{AB}, $\ker f = \operatorname{eq}(f, 0)$, when it exists. Dually, the cokernel is given by $\operatorname{coker} f = \operatorname{coeq}(f, 0_{AB})$.

Like all equalisers, the kernel is a regular monomorphism and so bears interpretation as a regular subobject. Dually, cokernels are regular quotient objects. Moreover, as with all limits, the object itself is only unique up to isomorphism, but should we include the appropriate morphism, then the collection is unique up to unique isomorphism. Partly motivated by this, and partly by the ever-present weight of brevity, we use $\ker f$ to refer to both the object and the morphism $k : \ker f \to \operatorname{dom} f$ wherever context would disambiguate such a choice.

Non-example (III) 3.3.6

In RING, the category of unital rings, there are no categorical kernels as there can be no zero morphisms.

Remark (III) 3.3.7. The usual subtlety about referring to limits is exacerbated in the case of (co)kernels wherein we may wish to discuss objects such as $\ker \operatorname{coker} f$. In general categories, there isn't a canonical $\operatorname{coker} f$ from which to construct $\ker \operatorname{coker} f$. As such, any proof we give for $\ker \operatorname{coker} f$ and related notions is to be carefully understood to hold for a presupposed and indeed arbitrary choice of $\operatorname{coker} f$, but not for all such objects *at once*. In particular, as long as $f \cong g$ with either a domain or a codomain isomorphism, we see that $\ker f \cong \ker g$ and $\operatorname{coker} f \cong \operatorname{coker} g$ whenever they exist.

Prop. (III) 3.3.8. *In a category with zero object, for any monomorphism $m : A \to B$, $(\ker m, k : \ker m \to A) \cong (0, 0_A)$.*

Proof. For any arrow $k : K \to A$ such that $mk = 0_{AB}k$ we must have $mk = 0_{AB}k = 0_{KB} = m0_{KA}$ and so $k = 0_{KA}$. By definition 0_{KA} factors uniquely through the zero object. ∎

Cor. (III) 3.3.9. *If the category contains a zero object, then for all morphisms $f : A \to B$ with kernel and $g : C \to D$ with cokernel, $\ker \ker f \cong 0$ and $\operatorname{coker} \operatorname{coker} g \cong 0$.*

The reader may at this point be wondering about the reverse implication omitted from prop. (III) 3.3.8. As it turns out, it is not true in a *general* category. Indeed, we will need to first introduce the notion of abelian categories in order to satisfactorily demonstrate a sufficient condition.

Finally, we exhibit properties that we may intuitively suspect hold, based perhaps upon our experience with abelian groups.

Prop. (III) 3.3.10. *If the category contains a zero object, then the kernel of $0_{A,B} : A \to B$ is isomorphic to A.*

Proof. Observe that $0_{A,B} \operatorname{id}_A = 0_{A,B}$ and so we have a unique $u : A \to \ker 0_{A,B}$ with $ku = \operatorname{id}_A$. Then a simple universal property argument shows that $uk = \operatorname{id}_{\ker 0_{A,B}}$ so that $A \cong \ker 0_{A,B}$. ∎

Prop. (III) 3.3.11. *In a semi-additive category, if $(A = A_0 \oplus A_1, \iota, \pi)$ is a biproduct, then*

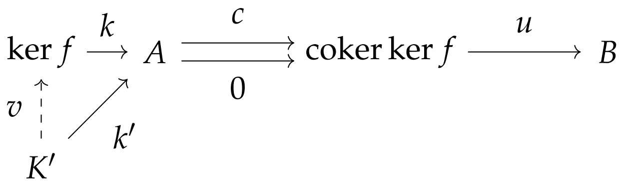
$$\iota_i \cong \ker \pi_j, \quad \pi_i \cong \operatorname{coker} \iota_j \quad (i \neq j)$$

Proof. Recall that the projections and inclusions satisfy $\pi_i \iota_j = \delta_{ij}$. We will show only that (A_i, π_i) is the coequaliser of $(0_{j,A}, \iota_j)$ for $i \neq j$ as the other proof is entirely similar.

First consider that we already have $\pi_i \iota_j = 0_{ij}$ and so we must only show that π_i is universal with respect to the coequaliser property. To that end, let $f : A \to B$ be an arrow with $f0_{j,A} = f\iota_j$. To see that there is a $u : A_i \to B$ with $u\pi_i = f$, recall that (prop. (III) 3.2.19) $f = \sum f\iota_i\pi_i$ but in this case, $f\iota_j = 0$ so that $f = f\iota_i\pi_i$ allowing us to write $u = f\iota_i$. That u is unique follows trivially from this, as if $v : A_i \to B$ had $v\pi = f$ then $v = v\pi_i\iota_i = f\iota_i = u$. ∎

Prop. (III) 3.3.12. *For arbitrary arrow f, whenever the appropriate objects exist, the following isomorphisms hold:* $\ker \operatorname{coker} \ker f \cong \ker f$ *and* $\operatorname{coker} \ker \operatorname{coker} f \cong \operatorname{coker} f$.

Proof. We show only the first as the second follows via dualisation. Let $f : A \to B$ and suppose $k : \ker f \to A$ and $c : A \to \operatorname{coker} \ker f$ exist. Observe that $fk = 0_{\ker f B} = f0_{\ker f A}$ and so f factors as $f = cu$ for unique $u : \operatorname{coker} \ker f \to B$. Now consider the following diagram



If $ck' = 0_{A \operatorname{coker} \ker f} k'$ then $fk' = uck' = u0_{A \operatorname{coker} \ker f} k' = 0_{AB} k'$ and so we have a unique $v : \ker f \to K'$ for which $kv = k'$, and thus $\ker \operatorname{coker} \ker f \cong \ker f$. ■

Cor. (III) 3.3.13. *If every arrow has a kernel and cokernel, then $f : A \to B$ is a kernel iff.* $f \cong \ker \operatorname{coker} f$.

3.4. Abelian categories

Now that we have seen how enrichment of CMon is variously equivalent to the presence of specific attributes of the category, and in particular how it lends itself to the powerful notion of a biproduct, we may be tempted to exchange CMon for a category with slightly more structure and reexamine the theory.

The theory of such structured categories is both rich and deep, but the direction of the document and brevity of the allocated time period for the completion of the work have conspired to constrain discussions to topical matters. Ergo, what follows is a brief outline of some surface results and elementary definitions in this direction, the sum total of which will set the stage for discussions in the next section.

Def. (III) 3.4.1. An Ab-enriched category is a category enriched over the symmetric monoidal category Ab of abelian groups, with the tensor product as that of $\mathbb{Z}$-modules.

In order to understand what such a category represents, we must carefully examine def. (III) 3.1.1 with the knowledge that morphism objects are now abelian groups. In doing so, we see that that composition must be an abelian group homomorphism, as it is an arrow in Ab. As such, we have the curious property that composition must be bilinear with respect to the $\mathbb{Z}$-module structure of the groups and the associated tensor product. In particular then, the category has zero morphisms and we are cured of one of the ailments of CMon enrichment.

Surprise (III) 3.4.2

We already know that enriching a singleton set yields a monoid object in the underlying monoidal category. Thus, we may be led to ask, by way of considering the simplest non-trivial Ab-enriched category, what is a monoid object in Ab?

A monoid object in Ab is an abelian group G together with a multiplication morphism $\mu : G \otimes G \to G$ and an identity morphism $\eta : \mathbb{Z} \to G$ satisfying the requisite relations of unitality and associativity. Moreover, the multiplication morphism must be bilinear (it is an arrow in Ab), and thus multiplication is distributive over addition. The careful reader will be quick to note that this means that we have simply arrived at the definition of a ring!

For this reason, Ab-enriched categories are sometimes referred to as *ringoids* as they represent the 'horizontal' categorical generalisation of rings.

Joke (III) 3.4.3. A ring is a ringoid with one object.

Remark (III) 3.4.4. Observe that Abelian groups are, in particular, commutative algebraic monoids and so every Ab-enriched category is also CMon-enriched. Thus, the theory established in section 3.3.2 applies here.

Example (III) 3.4.5

Ab is a closed symmetric monoidal category and so is enriched over itself, as the canonical example of an Ab-enriched category.

Remark (III) 3.4.6. Before we proceed to some results concerning Ab-enriched categories and their more structured brethren, we pause here to note that we already have a understanding of what functors between Ab-enriched categories should be. That is, we need only examine def. (III) 3.1.5 to find that such functors are morphisms of abelian groups which respect composition.

Prop. (III) 3.4.7. *In an* Ab*-enriched category, for a pair of parallel arrows $f, g : A \rightrightarrows B$, the following conditions are equivalent and the corresponding objects are isomorphic when they exist,*

1. $\mathrm{eq}(f,g)$ *exists*
2. $\ker(f-g)$ *exists*
3. $\ker(g-f)$ *exists*

Proof. Given that $\mathrm{eq}(f,g) = \mathrm{eq}(g,f)$ it suffices to show that $(1) \iff (2)$, for example. To that end, assuming (1) where $(E,e) = \mathrm{eq}(f,g)$ we posit an arrow $h : C \to A$ such that $(f-g)h = 0_{AB}$. However, $(f-g)h = 0_{AB} \iff fg = fh$ which gives a unique arrow $u : E \to C$ by the equaliser property with $e = uh$. The reverse implication and the rest of the proof proceed simply. ∎

Though this statement and its dual may be pleasing, simply enriching over Ab instead of over CMon does not bring us relevantly new, interesting results. The reader may perhaps convince himself that this is not surprising as, for example, biproducts only emerged from CMon-enrichment in the presence of finite products and a zero object. With this situation in mind, we introduce the following notion.

Def. (III) 3.4.8. An additive category is an Ab-enriched category with all finite products.

Given the contents of prop. (III) 3.2.24 we see that we may equally well have defined additive categories as Ab-enriched categories with all finite coproducts or biproducts.

If for no other reason than semantic similarity, the reader may wonder what relation additive categories have to *semi*-additive categories. Such a reader is to be congratulated for his directed questions, for they lead us to consider

Prop. (III) 3.4.9. *Any semi-additive category wherein the canonical enrichment over* CMon *extends a commutative group structure to the sets $\mathfrak{C}(A,B)$, is additive.*

Proof. We already know that semi-additive categories all finite biproducts, and so all we must demonstrate is that if the sets $\mathfrak{C}(A,B)$ have additive structures, then we have Ab enrichment.

This is almost completely trivial, however, as we already know that composition is distributive, associative, and unital in the proper ways (prop. (III) 3.2.17) and it is easy to see that positing the existence of additive inverses does not change any of this. Consequently, in order to prove the statement we really need only show that composition is a $\mathbb{Z}$-module morphism $\mathfrak{C}(B,C) \otimes_{\mathbb{Z}} \mathfrak{C}(A,B) \to \mathfrak{C}(A,C)$.

Thus, we aim to prove that for composable arrows $f : A \to B$, $g : B \to C$ we have $(ng)f = g(nf)$ for $n \in \mathbb{Z}$, where negative values of n are understood to have the meaning $ng = |n|(-g)$. To this end, consider that for $n = 0$ the statement has already been proven (prop. (III) 3.2.17), and for positive n, $(ng)f = (\sum^n g)f = \sum^n gf = g(\sum^n f) = ngf$ by distributivity of composition, while the proof for negative n follows, *mutatis mutandis*. ■

Where before we had that the biproduct structure in semi-additive categories determined a unique bimonoid structure for every object and so a canonical enrichment over CMon, in additive categories we have the following stronger and appropriately more amazing result.

Prop. (III) 3.4.10. *In an additive category, any two additive structures on the same morphism set are necessarily isomorphic.*

Proof. The proof proceeds through the following steps. We first show that for a given biproduct $A \oplus A$, $\delta = \iota_0 - \iota_1 \cong \ker \nabla_A$. That is, the difference of the inclusion maps is determined by the limit and colimit structures of the category (biproducts and equalisers), up to isomorphism. Then we show that every difference of parallel arrows admits a unique decomposition in terms of δ. Thus, $f - g$ is determined by the very same structure. Finally, we note that $f + g = f - (0 - g)$ and so the entire additive structure on the morphism sets is determined, up to isomorphism, by the limit and colimit structures of the category.

To begin then, recall that for fixed A, the unique arrow $\nabla_A : A \oplus A \to A$ is determined by the universal property $\nabla_A \iota = \mathrm{id}_A$ and, by the biproduct property we have $\nabla_A = \pi_0 + \pi_1$. Now, let $\delta = \iota_0 - \iota_1$ and observe that $\nabla_A \delta = \mathrm{id}_A - \mathrm{id}_A = 0_{A,A}$. With this, we will show that $(A, \delta) \cong \ker \nabla_A$.

$$\begin{array}{ccccc} A & \xrightarrow{\delta} & A \oplus A & \overset{\nabla_A}{\underset{0}{\rightrightarrows}} & A \\ {\scriptstyle u}\uparrow & \nearrow {\scriptstyle f} & & & \\ B & & & & \end{array}$$

We have already seen that $\nabla_A \delta = 0$, so suppose there was an arrow $f : B \to A$ with $\nabla_A f = 0$, thereby enforcing $\pi_0 f + \pi_1 f = 0$. We wish to show that there exists a unique $u : B \to A$ such that the above diagram commutes. If we let $u = \pi_0 f$ then we have $\delta u = \iota_0 \pi_0 f - \iota_1 \pi_0 f$. However, $\pi_0 f = -\pi_1 f$ by assumption so that $\delta u = \sum \iota_i \pi_i f = f$ by prop. (III) 3.2.19. Further, suppose $v : B \to A$ had $\delta v = f$. Then $\iota_0 v = f + \iota_1 v$ and so $\pi_0 \iota_0 v = \pi_0 f + \iota_1 v$, ergo $v = \pi_0 f = u$.

Now for arbitrary parallel arrows $f, g : A \rightrightarrows B$, the biproduct structure on A allows us to give a unique arrow $[f, g] : A \oplus A \to B$ satisfying the universal properties $[f, g]\iota_0 = f$ and $[f, g]\iota_1 = g$. As such, it is simple to see that $f - g = [f, g]\delta$ and so the difference of parallel arrows is determined by δ. ■

Cor. (III) 3.4.11. *Let $\mathfrak{C}$ be an additive category, then by prop. (III) 3.2.17 $\mathfrak{C}$ is canonically enriched over* CMon. *If all the morphism sets additionally are commutative groups, then the additive structure is isomorphic to the original additive structure.*

With an elementary understanding of additive categories achieved, we briefly mention here functors between additive categories.

Def. (III) 3.4.12. A functor between additive categories is termed additive when it is an abelian group homomorphism on each morphism collection.

Happily, we have that additive functors automatically respect biproducts.

Prop. (III) 3.4.13. *A functor between additive categories is additive iff. it preserves finite biproducts.*

Proof. Recall that a biproduct (def. (III) 3.2.23) was given determined entirely by its projections, inclusions and the equations relating them. In particular, $\pi_i \iota_j = \delta_{ij}$ and $\sum \iota_i \pi_i = \mathrm{id}$. Given that each equation is preserved by an additive functor, so too are biproducts.

Conversely, suppose $F : \mathfrak{C} \to \mathfrak{D}$ preserves biproducts and consider parallel arrows $f, g : A \rightrightarrows B$. We will aim to show the middle equality in the following, thereby proving the result using prop. (III) 3.2.25.

$$F(f+g) = F(\nabla_B (f \oplus g) \Delta_A) = \nabla_{FB}(Ff \oplus Fg)\Delta_{FA} = Ff + Fg$$

We have isomorphisms $\alpha : F(A \oplus A) \to FA \oplus FA$ and $\beta : F(B \oplus B) \to FB \oplus FB$ which satisfy the properties $\pi_{FA_0}\alpha = F\pi_{A_0}$, $\alpha^{-1}\iota_{FA_0} = F\iota_{A_0}$, *etc.*, by definition. In particular then,

$$\pi_{FB_0}(Ff \oplus Fg) = Ff\pi_{FA_0} = FfF\pi_{A_0}\alpha^{-1} = F\pi_{B_0}F(f \oplus g)\alpha^{-1} = \pi_{FB_0}\beta F(f \oplus g)\alpha^{-1}$$

and similarly for the other projection, π_{FB_1}. Thus, $F(f \oplus g) = \beta^{-1}(Ff \oplus Fg)\alpha$ by universal property. Furthermore, $\pi_{FA_0}\Delta_{FA} = \mathrm{id}_{FA} = F\pi_{A_0}F\Delta_A = \pi_{FA_0}\alpha F\Delta_A$ and so by universal property, $F\Delta_A = \alpha^{-1}\Delta_{FA}$. Dually, $F\nabla_B = \nabla_{FB}\beta$ and the result follows. ∎

Cor. (III) 3.4.14. *A functor is additive iff. it preserves finite biproducts iff. it preserves finite products iff. it preserves finite coproducts.*

Proof. We already have that a functor is additive iff. it preserves finite biproducts and so it remains to be shown that preserving finite biproducts is equivalent to preserving finite products (and by duality, finite coproducts). However, due to prop. (III) 3.2.24, this follows trivially. ∎

Now that we are satisfied with some of the basic matter concerning semi-additive and additive categories, it is time to introduce yet more structure.

Def. (III) 3.4.15. An additive category is said to be pre-abelian if every arrow has a kernel and cokernel.

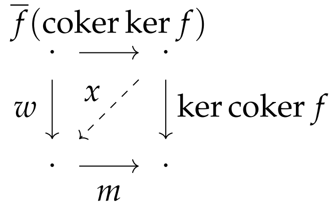
Non-example (III) 3.4.16

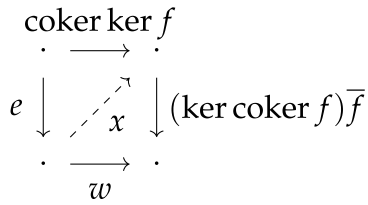
> The category of free abelian groups is not pre-abelian as there are no cokernels in general.

We are quick to note that pre-abelian categories are, in particular, finitely complete and cocomplete (as they have all (co)equalisers via prop. (III) 3.4.7 and (co)products by definition) and may have, in fact, been equivalently defined as Ab-enriched categories which are finitely complete and cocomplete.

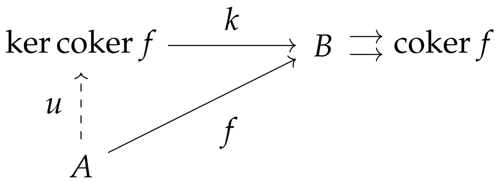
Prop. (III) 3.4.17. *In a pre-abelian category,*

1. *Every arrow admits a canonical factorisation as* $f = (\ker \operatorname{coker} f)\overline{f}(\operatorname{coker} \ker f)$
2. *If* $f = mw$ *where* m *is a kernel, then there is a unique monomorphism* x *such that the below-left diagram commutes. Dually, if* $f = we$ *where* e *is a cokernel then there is a unique epimorphism* x *such that the below-right diagram commutes*



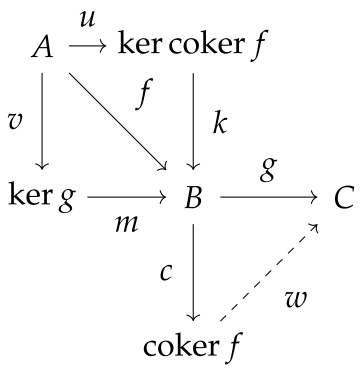


Proof. First we write $f = (\ker \operatorname{coker} f)u$ through the universal property of $\ker \operatorname{coker} f$,



Note that $f(\ker f) = 0$, but $f = (\ker \operatorname{coker} f)u$ with $\ker \operatorname{coker} f$ monic so $u(\ker f) = 0$ and we may factor $u = \overline{f} \operatorname{coker} \ker f$ by the coequaliser property.

For (2), suppose that $f = mv$ where $m = \ker g$ for some arrows $v : A \to \ker g$, $m : \ker g \to B$, and $g : B \to C$ and consider the below diagram.



By assumption the top square commutes and we have $gm = 0$. As such, $gf = gmv = 0$ and so by the cokernel property of $c = \operatorname{coker} f$ we have a unique $w : \operatorname{coker} f \to C$ for which $g = wc$. Observe then that $gk = wck = w0 = 0$ as $k = \ker c$ and so by the kernel property of $m = \ker g$ we must have a unique $x : \ker \operatorname{coker} f \to \ker g$ with $k = mx$ (which is easily seen to make x a monomorphism). Furthermore, $mv = f = ku = mxu$ and as m is a monomorphism by assumption we have $v = xu$. Dualisation of this argument completes the proof. ■

Regrettably, $\overline{f}$ is not monic or epic in general pre-abelian categories. In the same way that examining the restriction that the canonical morphism from coproducts to products be an isomorphism led to categories with interesting structure, we desire conditions for $\overline{f}$ to be an isomorphism.

Prop. (III) 3.4.18. *In a pre-abelian category, if every monomorphism is a kernel and every epimorphism is a cokernel, then for every arrow f, then canonical arrow $\overline{f} : \operatorname{coker}\ker f \to \ker\operatorname{coker} f$ is an isomorphism.*

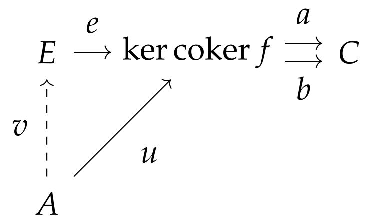
Before we prove this, we first give a minor technical result.

Lem. (III) 3.4.19. *In a pre-abelian category, if every monomorphism is a kernel and every epimorphism is a cokernel then any morphism which is monic and epic is an isomorphism.*

Proof. Assume f is monic and epic. As f is monic, $f = \ker g$ for some arrow g. However, $gf = 0$ gives $g = 0$ as f is an epimorphism. Thus, by prop. (III) 3.3.10, $f = \ker g \cong \mathrm{id}$. ∎

Proof (prop. (III) 3.4.18). We show that $\overline{f}$ is both epic and monic, and consequently an isomorphism given the assumptions.

Let $u = \overline{f}\operatorname{coker}\ker f$ and consider a pair of parallel arrows $a, b : \ker\operatorname{coker} f \rightrightarrows C$ such that $au = bu$ and take their equaliser, as in the following diagram.



Recall that $f = ku = kev$ and so $m = ke$ is a monomorphism which has $f = mv$. By assumption, m is a kernel and so we may apply prop. (III) 3.4.17 to find a unique monomorphism x with $k = mx = kex$. Thus $ex = \mathrm{id}$ and so $ae = be \implies a = b$, and from $u = \overline{f}\operatorname{coker}\ker f$ being epic it easily follows that $\overline{f}$ is too. Similarly, we perform the dual of the above proof to $v = (\ker\operatorname{coker} f)\overline{f}$ to find that v is a monomorphism and so $\overline{f}$ is both monic and epic. Thus, $\overline{f}$ is an isomorphism (lem. (III) 3.4.19). ∎

The observant reader will note that in a rather elementary manner, the above conditions are also necessary.

Prop. (III) 3.4.20. *In a pre-abelian category, every monomorphism is a kernel and every epimorphism is a cokernel iff. for every arrow the canonical arrow $\overline{f} : \operatorname{coker}\ker f \to \ker\operatorname{coker} f$ is an isomorphism.*

Proof. We have already shown the 'only if' part. Assume that m is a monomorphism and write $m = (\ker\operatorname{coker} m)\overline{m}(\operatorname{coker}\ker m)$ by prop. (III) 3.4.17, for $\overline{m}$ an isomorphism. By prop. (III) 3.3.8, $\ker m \cong 0$ and so $\operatorname{coker}\ker m \cong \operatorname{coker} 0 \cong \mathrm{id}$ (prop. (III) 3.3.10), making $m \cong \ker(\operatorname{coker} m)$ and thus a kernel. By dualisation, $e \cong \operatorname{coker}(\ker e)$ for e an epimorphism. ∎

Cor. (III) 3.4.21. *In a pre-abelian category, where every monomorphism is a kernel and every epimorphism is a cokernel,* $\ker f \cong 0 \iff f$ *is a monomorphism.*

Proof. The 'only if' is given by prop. (III) 3.3.10 and the 'if' by the above proof. ■

With the sufficiency and necessity of the condition achieved, we may question the extent to which such a factorisation is unique. In order to answer this, we must transition to a setting where such a factorisation always exists.

Def. (III) 3.4.22. A pre-abelian category wherein every monomorphism is a kernel and every epimorphism is a cokernel is said to be abelian.

It is no accident of naming that we have chosen the adjective abelian. Indeed,

Example (III) 3.4.23

> Ab is an abelian category. It is Ab-enriched, it supports finite biproducts in the usual manner, every arrow has a kernel and cokernel (again in the usual manner), and every monomorphism is a kernel ($G \to H \rightrightarrows H/\operatorname{im}$) and every epimorphism is a cokernel.

Returning to the matter of factorisation – as it happens, not only is the canonical factorisation unique in abelian categories, but there is a far more general result which implies it.

Prop. (III) 3.4.24. *In an abelian category, for every commutative square of arrows* $bf = f'a$ *(below left), if we write* $f = me$ *for* $m = (\ker \operatorname{coker} f)\overline{f}$ *and* $e = \operatorname{coker} \ker f$ *and* $f' = m'e'$ *for* m' *monic and* e' *epic, then there exists a unique* u *such that the diagram below right commutes*

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle a}\downarrow & & \downarrow{\scriptstyle b} \\ A' & \xrightarrow[f']{} & B' \end{array} \qquad \begin{array}{ccccc} A & \xrightarrow{e} & \operatorname{coker}\ker f & \xrightarrow{m} & B \\ {\scriptstyle a}\downarrow & & \downarrow{\scriptstyle u} & & \downarrow{\scriptstyle b} \\ A' & \xrightarrow[e']{} & \cdot & \xrightarrow[m']{} & B' \end{array}$$

Proof. Let $k = \ker f$ so that $ek = 0$. Consequently, $mek = 0$ and $fk = 0$ and $bfk = 0$ and $f'ak = m'e'a = 0$ so that $e'a = 0$. As such, $e'a$ factors uniquely through $\operatorname{coker}\ker f$ as $ue = e'a$. Finally, $m'ue = f'a = bf = bme$ implies that $m'u = bm$. ■

Cor. (III) 3.4.25. *In an abelian category, a mono-epi factorisation of an arrow* $f = me$ *is unique up to isomorphism.*

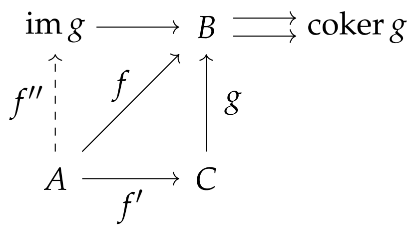
Proof. Write $f = me = m'e'$ and apply prop. (III) 3.4.24 to the degenerate square to find a unique u with $e' = ue$ and $m = m'u$ making u both epic and monic and so an isomorphism (lem. (III) 3.4.19). ■

Def. (III) 3.4.26. In an abelian category, if we write $f = (\ker\operatorname{coker} f)\overline{f}(\operatorname{coker}\ker f)$ then we say that $\operatorname{im} f = \ker\operatorname{coker} f$ and $\operatorname{coim} = \operatorname{coker}\ker f$.

Remark (III) 3.4.27. This definition is well chosen indeed. First, we have that $\operatorname{im} m \cong m$ for monomorphism m and the dual result. To see this, combine the fact that every monomorphism is a kernel with prop. (III) 3.3.12. Second, a rephrasing of prop. (III) 3.4.20 would be the statement that abelian categories are precisely the pre-abelian categories wherein the first isomorphism theorem holds ($\operatorname{im} \cong \operatorname{coim}$).

Moreover, abelian categories grant a convenient notion of quotient object more readily recognisable then merely an isomorphism class of epimorphisms.

Def. (III) 3.4.28. In an abelian category, if an arrow f factors through an arrow g, then we write g/f for $\operatorname{coker}(f'' : A \to \operatorname{im} g)$ where f'' is the unique arrow arising out of the below-right commutative diagram.



The dual construction, where f and g have common domain A and $f = f'g : A \to B$, gives rise to $\ker(f'' : B \to \operatorname{coim} g)$ and we abuse notation to write $g\backslash f$ for this case.

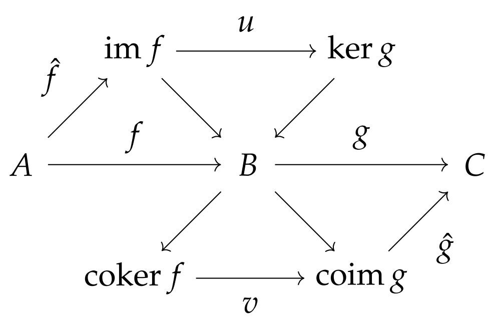
Remark (III) 3.4.29. Again, this coincides with what we might expect in Ab. The fact that f factors through g means that $\operatorname{im} f \subseteq \operatorname{im} g$ and so we can factor f through $A \to \operatorname{im} g \subseteq B$ via f'' which acts on elements as $f''(a) \mapsto f(a)$ – essentially a codomain restriction of f. Then, with the interpretation that $\operatorname{coker} f'' = \operatorname{im} g / \operatorname{im} f''$ we see that we have indeed created a reasonable definition of g/f. Moreover, if $C = B$ and $g = \mathrm{id}_C$ so that $f = f'$ then we see that $\operatorname{im} g = \mathrm{id}_C$ so that $f'' = f$ and $g/f = C/\operatorname{im} f$, exactly as we would have liked.

With an eye to later sections, we consider the following statement.

Prop. (III) 3.4.30. *In an abelian category, if $gf = 0$ for composable arrows f and g, then the following are all isomorphic*

1. $\ker g / \operatorname{im} f$
2. $\operatorname{coim} g \backslash \operatorname{coker} f$
3. $\operatorname{im}(\ker g \to \operatorname{coker} f)$
4. $\operatorname{coker}(\operatorname{im} f \to \ker g)$
5. $\ker(\operatorname{coker} f \to \operatorname{coim} g)$
6. $\operatorname{coim}(\ker g \to \operatorname{coker} f)$

Proof. To begin then, let $f : A \to B$ and $g : B \to C$ be such that $gf = 0$. Using the canonical decomposition, write $f = (\operatorname{im} f)\hat{f}$ and $g = \hat{g}(\operatorname{coim} g)$ for $\hat{f}$ epic and $\hat{g}$ monic. Now, noting that $gf = 0 \implies g(\operatorname{im} f)\hat{f} = 0$ and $\hat{f}$ epic, factor $\operatorname{im} f$ through $\ker g$ and likewise $\operatorname{coker} f$ through $\operatorname{coim} g$ to arrive at the following commutative diagram.



Unlabelled, left to right, top to bottom, are $\operatorname{im} f$, $\ker g$, $\operatorname{coker} f$, and $\operatorname{coim} g$. With that achieved, consider that $\ker g / \operatorname{im} f = \operatorname{coker}(\operatorname{im} f \to \operatorname{im} \ker g)$ but $\operatorname{im} \ker g \cong \ker g$ (prop. (III) 3.3.12). Thus $\ker g / \operatorname{im} f \cong \operatorname{coker} u$ and dually $\operatorname{coim} g \backslash \operatorname{coker} f \cong \ker v$, giving (1)≅(4) and (2)≅(5). Then, writing λ for $\operatorname{coker} f \ker g$, by prop. (III) 3.4.20 we have that $\operatorname{coim} \lambda \cong \operatorname{im} \lambda$ thereby showing (3)≅(6).

Next, consider that $\lambda u = \operatorname{coker} f \ker g u = \operatorname{coker} f \operatorname{im} f = 0$ and similarly that $\operatorname{coker} f \ker g \ker \lambda = \lambda \ker \lambda = 0$ so that we may find $\ker \lambda \cong \ker(\operatorname{coker} f) = \operatorname{im} f$ with unique isomorphism $\mu : \operatorname{im} f \to \ker \lambda$ having $(\ker \lambda)\mu = u$. As such, it follows that $\operatorname{coker} u \cong \operatorname{coker}(\ker \lambda) = \operatorname{coim} \lambda$ giving (4)≅(6). Dualisation yields (3)≅(5) thereby completing the proof. ■

Remark (III) 3.4.31. In the above proof we showed that $\operatorname{im} f \cong \ker \lambda$, which is essentially the generalised version of the statement that $gf = 0$ forces $\operatorname{im} f \subseteq \ker g$, should we view matters in Ab and see λ as the composite $\ker g \subseteq B \twoheadrightarrow B / \operatorname{im} f$.

To conclude this section, we note that we may have instead defined abelian categories in terms of the existence of certain objects and *derived* the additive structure on the morphism collections in a manner reminiscent to that of semi-additive categories. In particular, it is a theorem that

Thm. (III) 3.4.32. *A category is abelian iff all of the following hold,*

1. *there is a zero object,*
2. *every pair of objects has a product and a coproduct,*
3. *every arrow has a kernel and a cokernel,*
4. *every monomorphism is a kernel; every epimorphism is a cokernel*

While interesting and certainly in the spirit of the exposition so far, this proof would require a few involved technical lemmas which would consume too much space-time. As such, the ever curious reader is directed to [Bor94] for a full and lucid exposition.

4. Homology

Now that we have established some of the basic definitions and elementary results concerning abelian categories, we may use this language as a platform to discuss exactness and homology functors, and ultimately to briefly phrase the classical singular homology in a more general fashion.

However, due once more to the extremely short time-frame permitted to the author, our treatment of these notions will be sparing and we shall introduce only the barest of definitions in an attempt to charge towards the statement of singular homology as economically as possible. That is to say, we shall not explore at all the elementary diagram lemmas nor indeed shall we investigate any of the theory concerning projective modules, Ext and Tor, or particular examples of chain homology beyond a simple outline of simplicial homology. Moreover, we shall omit many important statements concerning homology functors and their specific instances – statements such as the homotopy invariance of H_n which lend themselves to the greater context.

Nevertheless, we will strive to give – if only in the broadest of strokes – an idea of the foundational definitions, if not some discussion of a subset of the core objects of concern, so that further directed investigation may be made from here.

4.1. Exactness

Def. (III) 4.1.1. A pair of morphisms $f : A \to B$ and $g : B \to C$ are said to be exact if $\operatorname{im} f \cong \ker g$. An exact sequence in a category with zero morphisms is given by a sequence of objects (A_n) and accompanying morphisms $f_n : A_n \to A_{n+1}$ such that each pair (f_n, f_{n+1}) is exact.

Prop. (III) 4.1.2. *In an abelian category,* $\operatorname{im} f \cong \ker g \iff \operatorname{coker} f \cong \operatorname{coim} g$.

Proof. Recall that for every morphism $f = (\operatorname{im} f)\overline{f}(\operatorname{coim} f)$ and in particular $\operatorname{im} f = \ker \operatorname{coker} f$ and $\operatorname{coim} f = \operatorname{coker} \ker f$. As such, if $\ker \operatorname{coker} f \cong \ker g$ then $\operatorname{coker} f \cong \operatorname{coker} \ker \operatorname{coker} f \cong \operatorname{coker} \ker g = \operatorname{coim} g$ where the first isomorphism is due to prop. (III) 3.3.12. The other direction follows by dualisation. ■

Def. (III) 4.1.3. A short exact sequence is an exact sequence of the form

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

In many instances, we will attempt to avoid diagrams such as the above, where instead an 'arrows-only' version of the same sequence would suffice. That is, we write $f \to g$ where we mean instead $\cdot \xrightarrow{f} \cdot \xrightarrow{g} \cdot$.

Prop. (III) 4.1.4. *In an abelian category, an exact pair of morphisms $f \to g$ form a short exact sequence iff. both*

1. *f is monic, g is epic*
2. $f \cong \ker g$ *(equivalently,* $\operatorname{coker} f \cong g$*)*

The proof is not particularly enlightening, and relies on manipulations of the canonical decomposition of morphisms in abelian categories (props. (III) 3.4.17 and (III) 3.4.18) using a few properties of kernels and cokernels (cor. (III) 3.4.21 and props. (III) 3.3.10 and (III) 3.3.12). Nevertheless, it serves to show that we can manipulate objects in abelian categories as though they had many of the familiar properties of abelian groups.

Proof. Assume that $0 \to f \to g \to 0$ forms a short exact sequence, that is, $\operatorname{im} 0 \cong \ker f$, $\operatorname{im} f \cong \ker g$ and $\operatorname{im} g \cong \ker 0$. Then, $\operatorname{im} 0 = \ker \operatorname{coker} 0 \cong \ker \mathrm{id} \cong 0$ so that $0 \cong \ker f$ and thus f is monic. Dually, g must be epic. Furthermore, recalling that $f = (\operatorname{im} f)\overline{f}(\operatorname{coim} f)$ where $\operatorname{coim} f = \operatorname{coker} \ker f \cong \operatorname{coker} 0 \cong \mathrm{id}$ we have $f \cong \operatorname{im} f$ and so $f \cong \operatorname{im} f \cong \ker g$ by exactness. Dually we may have pursued the exactness property of $\operatorname{coker} f \cong \operatorname{coim} g$ to find $\operatorname{coker} f \cong g$.

Then, assume (1) and (2) to find that $\ker f \cong 0 \cong \operatorname{im} g$, and similarly $\operatorname{im} g = \ker \operatorname{coker} g \cong \ker 0$ so that $0 \to f$ and $g \to 0$ are exact. Moreover, $f \cong \ker g \implies \ker \operatorname{coker} f \cong \ker \operatorname{coker} \ker g \cong \ker g$ and so the sequence is short and exact. ■

Prop. (III) 4.1.5. *In an abelian category,* $\ker f \to \operatorname{coim} f$ *and* $\operatorname{im} f \to \operatorname{coker} f$ *form short exact sequences for any arrow* f.

Proof. Both $\ker f$ and $\operatorname{im} f$ are monomorphism, and both $\operatorname{coim} f$ and $\operatorname{coker} f$ are epimorphisms. As such, we must only check that $\ker f \cong \ker \operatorname{coim} f$ and $\operatorname{im} f \cong \ker \operatorname{coker} f$. The second is true by definition and the first admits a simple proof as $\ker \operatorname{im} f = \ker \operatorname{coker} \ker f \cong \ker f$ (prop. (III) 3.3.12). ■

As we can see, by assuming that the underlying category is abelian we are afforded some convenient reformulations of exactness and familiar results. Partly inspired by this, we shall restrict all further discussion in this section and those that follow, by implicitly assuming that whenever exactness or chain complexes arise the ambient category is abelian.

Given this context, and that we already know what additive functors are, we may be tempted to define 'abelian' functors which respect that key advantage of abelian categories over additive ones, viz., finite completeness and cocompleteness. To this end, we turn to finitely continuous, cocontinuous and bicontinuous functors.

Remark (III) 4.1.6. The current and popular terminology for finitely continuous, cocontinuous and bicontinuous functors (in the context of homological algebra, and somewhat beyond) is, respectively, left-exact, right-exact and exact. In an effort to remain standard in this matter we shall employ these terms.

As an immediate consequence, exact functors between abelian categories preserve exact sequences and so fulfil an important role in the study of such objects. Observe further that we did not define exact functors between abelian categories to be additive, but it is a consequence of cor. (III) 3.4.14 that left- and right- exact functors between abelian categories are additive. In fact,

Prop. (III) 4.1.7. *A functor between abelian categories is left-exact iff. it is additive and it preserves kernels.*

Proof. Combine cor. (III) 3.4.14 and prop. (III) 3.4.7 and that finite completeness is equivalent to the existence of all finite products and finite equalisers. ■

This allows us to give an alternate characterisation of exact functors.

Prop. (III) 4.1.8. *An additive functor between abelian categories is exact iff. it preserves short exact sequences.*

Proof. Combine props. (III) 4.1.4 and (III) 4.1.7. ■

More than this, exact functors play a crucial role in enabling discussion of exact sequences in general abelian categories. In particular, though we have not explored "diagram chasing", many proofs are made tractable by explicitly tracing an element about a diagram as it undergoes the actions of various morphisms. Naturally, such an approach is not possible in general abelian categories and so we must find an alternative. One such is the following.

Thm. (III) 4.1.9 (Freyd-Mitchell embedding)**.** *Every small abelian category admits a fully-faithful and exact functor to* R-Mod *for some unital ring* R*.*

Regrettably, the proof is well beyond our means to sketch. A full version with all the necessary exposition may be found in [Fre64].

Thus, whenever we need to prove a result concerning exactness or indeed any forms of kernels or images, we may simply trace the action of maps as though we were in R-Mod and the result would be valid, independent of whether there exists an appropriate notion of elements of objects in the abelian category in question.

While this is indeed convenient, it may be troublesome that the enabling theorem is so far beyond the scope of the content thus far. [ML97] provides for us an alternate view of the scenario by defining a general notion of "members" of an object in an abelian category and showing that such members, equivalence classes of maps with codomain of interest, behave just as though they were "elements" of the object in question, thereby obviating the need for such powerful and advanced considerations as the theorem of Freyd-Mitchell.

With all of this, we would be able to consider such statements as the five lemma, given below, which prove useful in more advanced considerations in homological algebra.

Lem. (III) 4.1.10 (Five lemma)**.** *Given the below commutative diagram in an abelian category, if the top row is exact, m and p are isomorphisms, l is an epimorphism, and q is a monomorphism, then n is an isomorphism.*

$$\begin{array}{ccccccccc} A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & D & \longrightarrow & E \\ \downarrow{\scriptstyle l} & & \downarrow{\scriptstyle m} & & \downarrow{\scriptstyle n} & & \downarrow{\scriptstyle p} & & \downarrow{\scriptstyle q} \\ A' & \longrightarrow & B' & \longrightarrow & C' & \longrightarrow & D' & \longrightarrow & E' \end{array}$$

4.2. Chain homology

Def. (III) 4.2.1. In an abelian category, a chain complex is a sequence of objects labelled by integers and composable arrows between them

$$\cdots \to C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \to \cdots$$

where $\partial_n \partial_{n+1} = 0$ for all n, oftentimes abbreviated as $(C_\bullet, \partial_\bullet)$ or $C_\bullet$.

We may be tempted to consider chains as objects all their own, and as such, we would require a suitable definition of morphisms between chain complexes. In what follows, we will write all maps within chains as ∂_n wherever unambiguous.

Def. (III) 4.2.2. A morphism of chains, $f_\bullet : C_\bullet \to D_\bullet$, is a collection of arrows $f_n : C_n \to D_n$ such the following diagram commutes.

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & C_{n+1} & \xrightarrow{\partial_{n+1}} & C_n & \xrightarrow{\partial_n} & C_{n-1} & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle f_{n+1}} & & \downarrow{\scriptstyle f_n} & & \downarrow{\scriptstyle f_{n-1}} & & \\
\cdots & \longrightarrow & D_{n+1} & \xrightarrow[\partial'_{n+1}]{} & D_n & \xrightarrow[\partial'_n]{} & D_{n-1} & \longrightarrow & \cdots
\end{array}$$

Remark (III) 4.2.3. Note that for the above diagram to commute, it is sufficient and necessary for each square to commute and so we really require that $\partial'_n f_n = f_{n-1}\partial_n$.

With that established, and the notion of composition of chain morphisms defined in the obvious manner, we write $\textsc{Ch}(\mathfrak{A})$ for the category of chain complexes over the abelian category $\mathfrak{A}$.

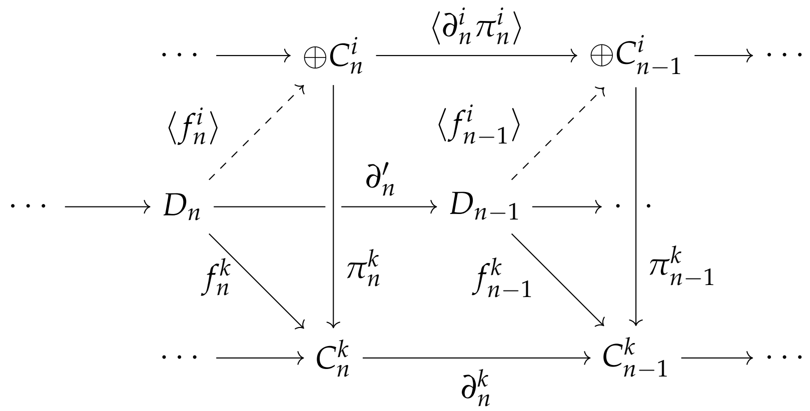
With the introduction of $\textsc{Ch}(\mathfrak{A})$, an entire line of inquiry becomes available and we may ask about the nature of $\textsc{Ch}(\mathfrak{A})$, and in particular, the extent to which the structure of $\mathfrak{A}$ effects it.

Prop. (III) 4.2.4. *If $\mathfrak{A}$ is abelian, then $\textsc{Ch}(\mathfrak{A})$ has all finite biproducts, kernels and cokernels, and they are computed 'degree-wise': $\oplus(C^i_\bullet, \partial^i) \cong (\oplus C^i_\bullet, \langle \partial^i_\bullet \pi^i_\bullet \rangle)$, etc.*

Proof. Recall that given a finite set I and a collection of chain complexes $(C^i_\bullet, \partial^i_\bullet)_{i \in I}$, $C^i_n \in \operatorname{Obj}\mathfrak{A}$ for every $i \in I$, $n \in \mathbb{Z}$ so that $\oplus_I C^i_n$ exists as an object in $\mathfrak{A}$. With this in hand, we show that $(\oplus C^i_\bullet, \langle \partial^i_\bullet \pi^i_\bullet \rangle)$ is a chain and supports the correct universal property to be $\prod(C^i_\bullet, \partial^i_\bullet)$. By dualisation it will follow that $\coprod(C^i_\bullet, \partial^i) \cong (\oplus C^i_\bullet, \langle \partial^i_\bullet \pi^i_\bullet \rangle)$ and thus we conclude the existence of biproducts in $\textsc{Ch}(\mathfrak{A})$. First, observe that

$$\pi^k_n \langle \partial^i_n \pi^i_n \rangle \langle \partial^i_{n+1} \pi^i_{n+1} \rangle = \partial^k_n \partial^k_{n+1} \pi^k_{n+1} = 0$$

so that by universal property $\langle \partial^i_n \pi^i_n \rangle \langle \partial^i_{n+1} \pi^i_{n+1} \rangle = 0$. Then, suppose there was a chain $(D_\bullet, \partial'_\bullet)$ with chain maps $f^i_\bullet : D_\bullet \to C^i_\bullet$, and consider the following diagram.

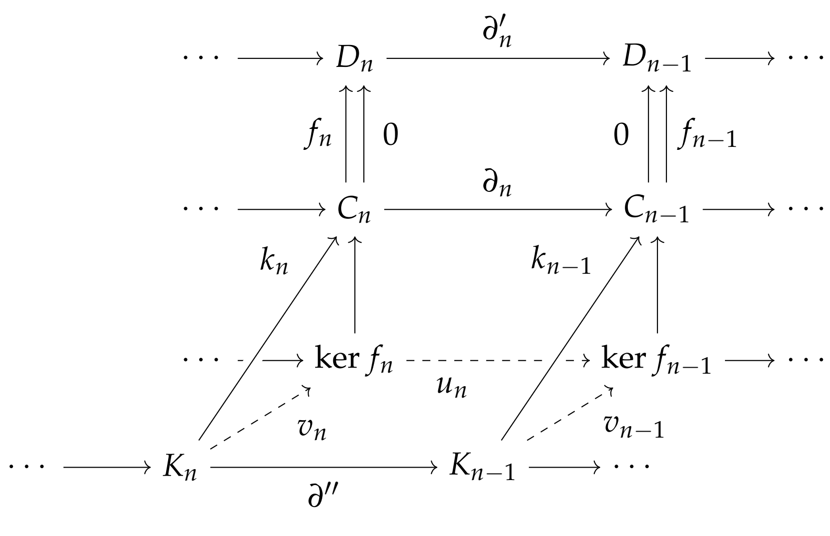


The bottom face commutes by assumption, and the left, right and front faces commute by definition of the maps $\langle f_n^i \rangle$, $\langle f_{n-1}^i \rangle$ and $\langle \partial_n^i \pi_n^i \rangle$ respectively. To see that the back face commutes, we must view the relevant composites as arrows $u_n : D_n \to \oplus C_{n-1}^i$ thereby uniquely characterising them by their projections. However,

$$\pi_{n-1}^k \langle \partial_n^i \pi_n^i \rangle \langle f_n^i \rangle = \partial_n^k f_n^k = f_{n-1}^k \partial_n' = \pi_{n-1}^k \langle f_{n-1}^i \rangle \partial_n'$$

The commutativity of the diagram thus follows and this, with dualisation, completes the proof of the existence of biproducts.

In an entirely similar vein, let $f_\bullet : C_\bullet \to D_\bullet$ be a chain morphism and suppose there was $k_\bullet : K_\bullet \to C_\bullet$ such that $f_n k_n = 0$ and consider the following diagram



Here we have noted that $f_{n-1}\partial_n \ker f_n = \partial_n' f_n \ker f_n = 0$ to find the unique arrow u_n with $(\ker f_{n-1})u_n = \partial_n \ker f_n$, and v_n, v_{n-1} arise via assumption. With that established, we note that the top face commutes by assumption, and the left and right triangles and the central square commute by definition of the maps v_n, v_{n-1} and u_n respectively. To see that the bottom face commutes, that $u_n v_n = v_{n-1}\partial_n''$, we make use of what is effectively a universal property argument.

Observe that $\ker f_{n-1}$ is a monomorphism and post-composition of both composites yields the same morphism so that the entire diagram commutes. Explicitly,

$$(\ker f_{n-1})u_n v_n = \partial_n (\ker f_n) v_n = \partial_n k_n = k_{n-1}\partial_n'' = (\ker f_{n-1}) v_{n-1} \partial_n''$$

All that remains concerning $\ker f_\bullet$ is to show that $u_n u_{n+1} = 0$ making $(\ker f_\bullet, u_\bullet)$ a chain. This matter is quickly laid to rest when we recall that $\ker f_{n-1}$ is a monomorphism, so that equality $(\ker f_{n-1})u_n u_{n+1} = \partial_n(\ker f_n)u_{n+1} = \partial_n\partial_{n+1}(\ker f_{n+1}) = 0 = (\ker f_{n-1})0$ gives the required statement. Dualisation completes the proof. ■

This result allows us to tie the proverbial knot and demonstrate that

Cor. (III) 4.2.5. $\mathrm{C{\scriptsize H}}(\mathfrak{A})$ *is abelian if* $\mathfrak{A}$ *is abelian.*

Proof. Under the evident, degree-wise additive structure $\mathrm{C{\scriptsize H}}(A)$ is certainly pre-abelian due to the previous result. In order for the category to be considered abelian, it remains to be shown, as per prop. (III) 3.4.20, that for a chain morphism $f_\bullet : C_\bullet \to D_\bullet$, $\operatorname{im} f_\bullet \cong \operatorname{coim} f_\bullet$.

To begin, consider the collections $(\operatorname{im} f_\bullet, \widehat{\partial}_\bullet)$ and $(\operatorname{coim} f_\bullet, \widetilde{\partial}_\bullet)$ where the maps $\widehat{\partial}_\bullet$ and $\widetilde{\partial}_\bullet$ arise out of the following commutative diagrams.

$$\begin{array}{ccccc} C_n & \rightrightarrows & D_n & \longrightarrow & \operatorname{coker} f_n \\ \downarrow{\scriptstyle \partial_n} & & \downarrow{\scriptstyle \partial'_n} & & \downarrow{\scriptstyle \partial''_n} \\ C_{n-1} & \rightrightarrows & D_{n-1} & \to & \operatorname{coker} f_{n-1} \end{array} \qquad \begin{array}{ccccc} \ker f_n & \longrightarrow & C_n & \rightrightarrows & D_n \\ \downarrow{\scriptstyle \partial'''_n} & & \downarrow{\scriptstyle \partial_n} & & \downarrow{\scriptstyle \partial'_n} \\ \ker f_{n-1} & \longrightarrow & C_{n-1} & \rightrightarrows & D_{n-1} \end{array}$$

$$\begin{array}{ccccc} \operatorname{im} f_n & \longrightarrow & D_n & \rightrightarrows & \operatorname{coker} f_n \\ \downarrow{\scriptstyle \widehat{\partial}_n} & & \downarrow{\scriptstyle \partial'_n} & & \downarrow{\scriptstyle \partial''_n} \\ \operatorname{im} f_{n-1} & \longrightarrow & D_{n-1} & \rightrightarrows & \operatorname{coker} f_{n-1} \end{array} \qquad \begin{array}{ccccc} \ker f_n & \rightrightarrows & C_n & \longrightarrow & \operatorname{coim} f_n \\ \downarrow{\scriptstyle \partial'''_n} & & \downarrow{\scriptstyle \partial_n} & & \downarrow{\scriptstyle \widetilde{\partial}_n} \\ \ker f_{n-1} & \rightrightarrows & C_{n-1} & \to & \operatorname{coim} f_{n-1} \end{array}$$

To see that these are chains, consider that $\operatorname{coim} f_{n+1}$ is an epimorphism so we check $\widetilde{\partial}_n\widetilde{\partial}_{n+1}\operatorname{coim} f_{n+1} = \widetilde{\partial}_n(\operatorname{coim} f_n)\partial_{n+1} = (\operatorname{coim} f_{n-1})\partial_n\partial_{n+1} = 0 = 0(\operatorname{coim} f_{n+1})$ and dually for $\widehat{\partial}_n\widehat{\partial}_{n+1} = 0$. With that established, we must show that the following diagram commutes in order demonstrate that the degree-wise isomorphism extends to a chain isomorphism $(\operatorname{im} f_\bullet, \widehat{\partial}_\bullet) \cong (\operatorname{coim} f_\bullet, \widetilde{\partial}_\bullet)$, where $\overline{f}_i$ is the usual isomorphism in $\mathfrak{A}$.

$$\begin{array}{ccc} \operatorname{im} f_n & \xrightarrow{\widehat{\partial}_n} & \operatorname{im} f_{n-1} \\ {\scriptstyle \overline{f}_n}\uparrow & & \uparrow{\scriptstyle \overline{f}_{n-1}} \\ \operatorname{coim} f_n & \xrightarrow[\widetilde{\partial}_n]{} & \operatorname{coim} f_{n-1} \end{array}$$

We shall show that the two composites are equal by post-composition with $\operatorname{im} f_{n-1}$, a monomorphism.

First observe that $(\operatorname{im} f_{n-1})\widehat{\partial}_n\overline{f}_n = \partial'_n(\operatorname{im} f_n)\overline{f}_n$ by the bottom-left diagram. For the second composite we pre-compose with $\operatorname{coim} f_n$ to find $(\operatorname{im} f_{n-1})\overline{f}_{n-1}\widetilde{\partial}_n(\operatorname{coim} f_n) = (\operatorname{im} f_{n-1})\overline{f}_{n-1}(\operatorname{coim} f_{n-1})\partial_n = f_{n-1}\partial_n = \partial'_n f_n = \partial'_n(\operatorname{im} f_n)\overline{f}_n(\operatorname{coim} f_n)$ so that we may infer $(\operatorname{im} f_{n-1})\overline{f}_{n-1}\widetilde{\partial}_n = \partial'_n(\operatorname{im} f_n)\overline{f}_n$, thereby concluding the proof. ■

Of course, there are now many questions which we may ask about $\textsc{Ch}(\mathfrak{A})$. For example, under what circumstances does iteration of $\textsc{Ch}$ produce genuinely new categories? If $\mathfrak{A}$ is monoidal, is $\textsc{Ch}(\mathfrak{A})$ monoidal too? Is it closed? Regrettably, such investigations would divert our attention too extensively to admit discussion here.

Despite our demanding that $\partial\partial = 0$, in general a chain complex need not also be an exact sequence considered as a diagram in the underlying category. In fact, by and large, homology is the study of the deviation from exactness of such complexes. In particular,

Def. (III) 4.2.6. Given a chain complex $C_\bullet$ we define the n-th homology object to be $H_n(C) = \ker \partial_n / \operatorname{im} \partial_{n+1}$, understood in the generalised sense of prop. (III) 3.4.30. If $H_n \cong 0$ then we say that the complex is exact in degree n.

It is a simple matter to check that our terminology of exactness is warranted.

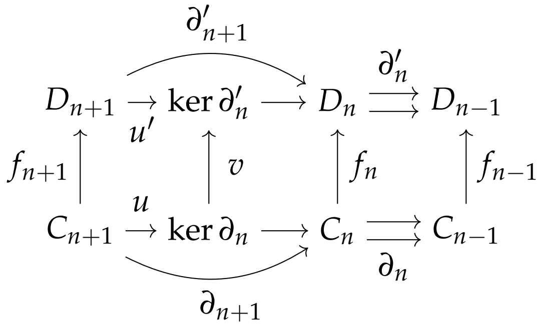
Prop. (III) 4.2.7. $H_n \cong 0 \implies \operatorname{im} \partial_{n+1} \cong \ker \partial_n$.

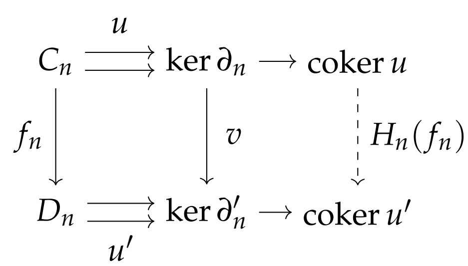
Proof. By prop. (III) 3.4.30, writing u for the unique arrow $\operatorname{im} \partial_{n+1} \to \ker \partial_n$, if $\operatorname{coker} u \cong 0$ then u must be epic (cor. (III) 3.4.21). However, by construction u is already monic and so u is an isomorphism (lem. (III) 3.4.19). ■

Moreover, should we carefully view H_n as an assignment of objects from $\textsc{Ch}(\mathfrak{A})$ to $\mathfrak{A}$, we may wonder whether H_n extends to a functor. Indeed,

Prop. (III) 4.2.8. *For each $n \in \mathbb{Z}$, $H_n : \textsc{Ch}(\mathfrak{A}) \to \mathfrak{A}$ is a functor.*

Proof. Let $f_\bullet : (C_\bullet, \partial_\bullet) \to (D_\bullet, \partial'_\bullet)$, be a morphism of chain complexes. Recall that $H_n(C) = \ker \partial_n / \operatorname{im} \partial_{n+1} \cong \operatorname{coker}(\operatorname{im} \partial_{n+1} \to \ker \partial_n)$ (prop. (III) 3.4.30). Further, if we write $\partial_{n+1} = (\operatorname{im} \partial_{n+1})\widehat{\partial_{n+1}}$ for $\widehat{\partial_{n+1}}$ epic then it is apparent that we have $H_n(C) \cong \operatorname{coker}(\operatorname{im} \partial_{n+1} \to \ker \partial_n) \cong \operatorname{coker}(C_{n+1} \to \ker \partial_n)$ and so we construct $H_n(f_n) : \operatorname{coker}(C_{n+1} \to \ker \partial_n) \to \operatorname{coker}(D_{n+1} \to \ker \partial'_n)$.





In the above-left diagram, we have observed that $\partial_n \partial_{n+1} = 0$ to retrieve the arrow u and similarly u'. Then we noted that $\partial' f_n \ker \partial_n = f_{n-1} \partial_n \ker \partial_n = 0$ to find the arrow v. By the universal property of $\ker \partial'_n$, it must be the case that $v = u' f_{n+1} u$ and so the diagram commutes. Then, turning to the above-right diagram, observe that $(\operatorname{coker} u')vu = (\operatorname{coker} u')u' f_{n+1}$ as the above-left diagram commutes, to find the unique arrow $H_n(f_n)$ as desired. Given that we have defined $H_n(f_n)$ by universal property, it is readily apparent that $H_n(g_n f_n) = H_n(g_n) H_n(f_n)$ and it is a simple matter to see that $H_n(\mathrm{id}_n) = \mathrm{id}_{\operatorname{coker} u}$. Finally, to conclude the proof recall that we simply have domain and codomain isomorphisms on $H_n(f)$ so as to interpret it as an arrow $H_n(C) \to H_n(D)$. ■

As has been indicated, many more topics and questions concerning H_n are immediately apparent, but we shall not have time to explore them.

4.3. The simplex category

The goal of this section is to introduce a means to discuss the classical notion of singular homology (with which some familiarity on the part of the reader is assumed) in a sufficiently general context. In particular, we do not wish to constrain our considerations to the case of classical singular homology – chains of free abelian groups generated by continuous functions from the standard topological simplexes $\{(t_0, \ldots, t_n) \in \mathbb{R}^{n+1} \mid \sum t_i = 1,\ t_i \geq 0\}$ into the space of concern with the morphisms of the chain given by the alternating sum of face maps. Indeed, as we shall see, we shall generalise the notion of standard simplex in such a way that we will be able to realise this construction in a categorical manner such that we may work in any abelian category.

We begin by introducing a particular monoidal category that is of central concern.

Def. (III) 4.3.1. The augmented simplex category, denoted Δ_a, is the category whose objects are finite ordinals and whose morphisms are order preserving functions. The simplex category is the full subcategory of non-zero ordinals and is denoted Δ.

Remark (III) 4.3.2. When we speak of ordinals in this section we will follow the standard definition due to Von Neumann which gives $0 = \phi$, $1 = \{0\}$, *etc.* It can be shown that each well-ordered set is order-isomorphic to precisely one such set and so, in this way, we have canonically chosen a representative of each equivalence class. Thus we will write $1 + 1 = 2$ with direct equality, and so on.

Recall that given two ordinals we may take their ordinal sum to arrive at a third ordinal, and that this sum is associative. Noting that there is an order isomorphism $n \cong \{\star\} \times n$, the ordinal sum $n + m$ is simply the usual coproduct of sets endowed with the evident total ordering that has $n < a$ for all $a \in m$. Consequently, $n = n + \phi = \phi + n$ and associativity of this operation can inductively be shown to be strict, where the base case is $(n + m) + \phi = n + (m + \phi)$. Moreover, $+$ extends to a map on morphisms in Δ_a as we define $f + f' : n + n' \to m + m'$ through

$$(f + f')(a) = \begin{cases} f(a), & a \in n \\ m + f'(a), & \text{otherwise} \end{cases}$$

It may be shown that this extension is functorial, making $+ : \Delta_a \times \Delta_a \to \Delta_a$ a bifunctor, thus rendering $(\Delta_a, 0, +)$ a strict monoidal category.

Def. (III) 4.3.3. Let $\delta^n_k : n \to n + 1$ be the injective order preserving function from n to $n + 1$ that omits $k \in n + 1$ in its image, $\delta_k(n) = \{0, \ldots, k - 1, k + 1, \ldots, n\} \subset n + 1$. Complementary to this, we write $\sigma^n_k : n + 1 \to n$ for the surjective order preserving function which does not increase on $k \in n + 1$, $\sigma^n_k(k) = \sigma^n_k(k + 1)$.

Remark (III) 4.3.4. Due to their geometric interpretation, the maps δ and σ are commonly referred to as the *coface* and *codegeneracy* maps.

Prop. (III) 4.3.5. *The following identities hold*

1. $j < k \implies \delta_j^{n+1}\delta_k^n = \delta_{k+1}^{n+1}\delta_j^n$

2. $j \le k \implies \sigma_j^{n-1}\sigma_k^n = \sigma_k^{n-1}\sigma_{j+1}^n$

3. $\sigma_j^n\delta_k^n = \begin{cases} \delta_k^{n-1}\sigma_{j-1}^{n-1}, & k < j \\ \delta_{k-1}^{n-1}\sigma_j^{n-1}, & k > j+1 \\ \mathrm{id}_n, & (k = j) \vee (k = j+1) \end{cases}$

Proof. Each statement may be shown via direct computation. ■

Enabled by the calculus outlined above, it is a theorem of [ML97] that every arrow in Δ_a admits canonical decomposition in terms of δ and σ. Our interest in this fact is limited to stating that functors whose domain is Δ_a are determined by the objects in their image and their action on δ and σ alone. The particulars of the result are as follows.

Thm. (III) 4.3.6 (Mac Lane). *Every arrow $f : n \to m$ in Δ_a admits a unique decomposition in terms of δ and σ as $f = \delta_{a_1} \circ \cdots \circ \delta_{a_k} \circ \sigma_{b_1} \circ \cdots \circ \sigma_{b_j}$ where $n + k = m + j$ and*

$$0 \le a_k < \cdots < a_1 < m, \quad 0 \le b_1 < \cdots < b_j < n - 1$$

Regrettably, that is the limit of our interest in Δ_a specifically, but the reader may rest assured that augmentations and related concepts have found employ in the general theory – we mention them only for completion as our true interest lies in Δ. With that established, we introduce some terminology.

Def. (III) 4.3.7. Given a category $\mathfrak{C}$, an augmented simplicial object in $\mathfrak{C}$ is a functor $S : \Delta_a^{\mathrm{op}} \to \mathfrak{C}$. Correspondingly, a simplicial object in $\mathfrak{C}$ is a functor $S : \Delta^{\mathrm{op}} \to \mathfrak{C}$. A morphism of simplicial objects is a natural transform between the functors.

Though obvious, we nevertheless make explicit the relations that d and s satisfy as a result of being the functorial images of δ^{op} and σ^{op}, for later reference.

Cor. (III) 4.3.8 (Dual to prop. (III) 4.3.5). *For a simplicial object $S : \Delta^{\mathrm{op}} \to \mathfrak{C}$, where $d = S\delta$ and $s = S\sigma$, the following identities hold.*

1. $j < k \implies d_j^{n-1}d_k^n = d_{k-1}^{n-1}d_j^n$

2. $j \le k \implies s_j^{n+1}s_k^n = s_{k+1}^{n+1}s_j^n$

3. $d_k^n s_j^n = \begin{cases} s_{j-1}^{n-1}d_k^{n-1}, & k < j \\ s_j^{n-1}d_{k-1}^{n-1}, & k > j+1 \\ \mathrm{id}_{Sn}, & (k = j) \vee (k = j+1) \end{cases}$

Example (III) 4.3.9

Simplicial sets are presheaves on Δ, and form the category sSet.

With this particular example, we may attempt to shed some light on the nature of simplicial objects via simplicial sets and some geometric allegories.

Recall that the Yoneda embedding gives an embedding $h_{-} : \mathfrak{C} \to [\mathfrak{C}^{\mathrm{op}}, \textsc{Set}]$ and so we may examine Δ under this embedding in $[\Delta^{\mathrm{op}}, \textsc{Set}] = \mathrm{s}\textsc{Set}$. In particular, let us consider the image of $n \in \Delta$ under this embedding and write Δ^n for $h_n = \Delta(-, n)$, which we shall term the *standard n-simplex*.

A convenient understanding of the standard n-simplex is as a generalised version of an ordered geometric simplicial complex. That is, we shall view Δ^n as a collection of sets (the images of the objects in Δ, no less) of geometric simplexes whose vertices are labelled by positive naturals ordered monotonically. These sets are indexed by the object of Δ in question (annoyingly this is one more than the *geometric* dimension, for two points make a line, *etc.*), $\Delta^n \sim \{S_1, S_2, \cdots\}$. However, we must also allow for 'degenerate' geometric simplexes in which some adjacent vertices coincide.

Explicitly, if we were to write out Δ^2 in this manner using the usual notation for simplexes on vertices, we would have sets $S_1 = \{[0], [1]\}$, $S_2 = \{[0,0], [0,1], [1,1]\}$, $S_3 = \{[0,0,0], [0,0,1], [0,1,1], [1,1,1]\}$, and so on – see fig. 3.1 for a visualisation. This notation is, obviously, doubly meaningful in this context. For Δ^n, the element $[v_1, \ldots, v_k] \in S_m$ with $k \in m$, $v_i \in n$, $i \leq j \implies v_i \leq v_j$, also records the order-preserving function in $\Delta(m, n)$. That is, we see $[v_0, \ldots, v_{m-1}]$ as the function which takes values $f(j) = v_j$.

Furthermore, this context allows for a convenient understanding of the face [3] and degeneracy maps. The face map is the image of δ_k under Δ^n and using our notation is an arrow $d_k = \Delta^n(\delta_k) : S_m \to S_{m-1}$ which takes a simplex and yields the embedded, 'lower dimensional' face by omitting the kth vertex, fig. 3.1. Really, however, this is just the usual precomposition of functions in $\Delta(m, n)$ by δ_k^{m-1}.

Dually, the degeneracy maps (images of σ under Δ^n) have type $S_m \to S_{m+1}$ and can be understood as taking a simplex $[v_0, \ldots, v_m]$ to the degenerate, 'higher dimensional' simplex $[v_0, \ldots, v_k, v_k, \ldots, v_m]$.

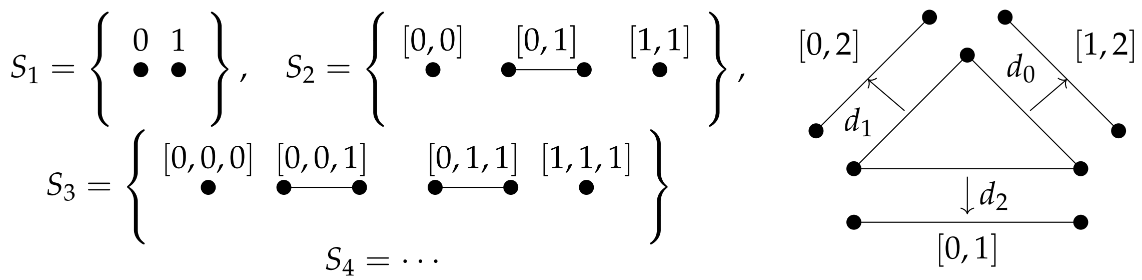


Figure 3.1: Left Δ^2 visualised, right face maps on $[0,1,2] \in S_3$ of Δ^3.

Though illuminating, as presented these notions only apply to the image of Δ embedded in $\mathrm{s}\textsc{Set}$. More general simplicial sets are unlike the standard n-simplexes described above (in the extreme, consider the constant simplicial set). However, they still obey the same relations due to their functorial nature and so may be *thought* of in the same manner. In fact, in general, the above ideas serve as as an effective mental model of the underlying nature of simplicial objects in general categories – though this must be used with caution, certainly since 'elements' do not always have analogues.

[3]Recall that we are talking about presheaves and so Δ^{op}, hence face instead of *co*face.

With the idea of simplicial objects established, we are now in a position to recall the classical singular homology and attempt to phrase it in a far more general manner.

Simplicial Homology

Let us write $|\Delta^n|$ for the standard topological n-simplex given by the convex hull $\{(t_0, \ldots, t_n) \in \mathbb{R}^{n+1} \mid \sum t_i = 1,\ t_i \geq 0\}$. Fixing a topological space X, it may be shown that the assignment $\mathcal{S} = \text{Top}(|\Delta^-|, X) : \Delta^{\text{op}} \to \text{Set}$ which acts in the evident manner on objects and has $S(\delta_k^{\text{op}})(f)$ as the restriction of f to the simplex without vertex k and likewise for sigma, is functorial and so forms a simplicial set. Thus, we have managed to phrase the selection of continuous maps from the standard geometric simplex into a space in terms of simplicial sets.

Remark (III) 4.3.10*.* That we have chosen to write $|\Delta^n|$ is no accident of notation. It is an indication that something far more interesting is happening behind the scenes. In general, there exists a way to construct a topological space from an arbitrary simplicial set, and moreover, this operation is functorial $|\cdot| : \text{sSet} \to \text{Top}$ and is termed the *realisation* functor. Unfortunately, we will not have time to study realisations and the wonderful concepts to which they lead.

Observe that, given a simplicial set $S : \Delta^{\text{op}} \to \text{Set}$, we may post-compose the free abelian group functor $Z : \text{Set} \to \text{Ab}$ to arrive at a simplicial group. Moreover, had these sets been arrived at via $\text{Top}(|\Delta^n|, X)$, we would have a complete categorical version of the classical singular homology construction. Thus, should we recast the classical alternating face map result in a suitably general manner, we will have succeeded in categorifying singular homology. Of course, there is no reason to restrict ourselves to free groups, nor indeed Ab. Thus, the general version of the chain part of the construction of singular homology would read as follows.

Prop. (III) 4.3.11. *Let A be a simplicial object in an abelian category $\mathfrak{A}$ with face maps d_k, and define $\partial_n : A_n \to A_{n-1}$ to be the map*

$$\partial_n = \sum_{k=0}^{n-1} (-1)^k d_k$$

then $(A_\bullet, \partial_\bullet)$ is a chain over $\mathfrak{A}$.

Proof. We must show that $\partial_n \partial_{n+1} = 0$, and this may be achieved by the usual direct expansion, noting well cor. (III) 4.3.8 (1) which enables the pairwise term cancellation. ∎

Thus, the categorical description of simplicial homology admits a neat summary.

Def. (III) 4.3.12. The simplicial homology of a simplicial object $A : \Delta^{\text{op}} \to \mathfrak{A}$, where $\mathfrak{A}$ is an abelian category, is the chain homology of the chain $(A_\bullet, \partial_\bullet)$ where $\partial_\bullet$ is as in prop. (III) 4.3.11.

5. Discussion

In summary then, we have made a careful study of the nature of categories supporting additional structure in order to arrive at the notion of abelian categories. Such categories were seen to have many of the desirable properties of categories such as R-Mod and Ab, and allowed us to redevelop and refine our intuitions about the notions of exactness and chains in greater generality.

With these notions in hand, we saw the ease with which classical singular homology became a simple example of a far broader class of related concepts. This class was effectively enabled by considering functors from the simplicial category. It is the opinion of the author that this short example serves as ample motivation (should the need exist) for the entire, categorical approach.

Regrettably, little of the exposition given bears any direct application to information manifolds beyond naïve computations of singular homology groups given the extant geometry – computations which certainly do not require the generality espoused. However, there is no doubt in the mind of the author, that without such carefully general statements enabling a thorough and 'example-agnostic' assessment of the state of the art, specific and novel results concerning information manifolds – should there be any within grasp – would certainly go unnoticed.

Thus, in the author's opinion, a far more thorough, far-reaching and detailed study of homological algebra will be absolutely necessary before any insightful applications of homological algebra to this particular case may be made.

Appendix

1. Adjunctions

Def. (A) 1.0.13. Given two categories $\mathfrak{C}$ and $\mathfrak{D}$, an adjunction between them is a pair of functors $L : \mathfrak{C} \to \mathfrak{D}$ and $R : \mathfrak{D} \to \mathfrak{C}$ such that for all objects $C \in \mathfrak{C}, D \in \mathfrak{D}$ there is an isomorphism $\mathfrak{D}(LC, D) \cong \mathfrak{C}(C, RD)$ which is natural in both arguments, that is, a natural isomorphism of morphism functors $\mathfrak{C}^{\text{op}} \times \mathfrak{D} \to \text{Set}$. We abbreviate this arrangement by writing $L \dashv R : \mathfrak{C} \to \mathfrak{D}$ and by stating that L is the left adjoint functor of R, or that R is the right adjoint functor of L.

Prop. (A) 1.0.14. *For functors $L \dashv R : \mathfrak{C} \to \mathfrak{D}$ the following are equivalent.*

1. $\mathfrak{D}(LC, D) \cong \mathfrak{C}(C, RD)$ *binaturally*
2. *There is a natural transformation $\eta : \text{id}_{\mathfrak{C}} \to RL$ called the unit and a natural transformation $\varepsilon : LR \to \text{id}_{\mathfrak{D}}$ called the counit such that the following diagrams commute (the triangle identities)*

$$\begin{array}{ccc} L & \xrightarrow{L\eta} & LRL \\ & \diagdown\!\!\diagdown & \downarrow{\scriptstyle \varepsilon L} \\ & & L \end{array} \qquad\qquad \begin{array}{ccc} R & \xrightarrow{\eta R} & RLR \\ & \diagdown\!\!\diagdown & \downarrow{\scriptstyle R\varepsilon} \\ & & R \end{array}$$

3. *For every $D \in \text{Obj}\,\mathfrak{D}$ there is a universal arrow (RD, ε) from L to D where R is the resultant functor. The dual statement is equivalent, too.*

Proof. This is a standard result, so we omit the proof. ∎

Prop. (A) 1.0.15. *Adjoints are unique up to isomorhpism.*

Proof. Suppose that $L \dashv R, R' : \mathfrak{C} \to \mathfrak{D}$, then in particular we have natural isomorphisms $\alpha : \mathfrak{C}(-, R-) \to \mathfrak{D}(L-, -)$ and $\beta : \mathfrak{D}(L-, -) \to \mathfrak{C}(-, R'-)$ and so a natural isomorphism $\beta\alpha : \mathfrak{C}(-, R-) \to \mathfrak{C}(-, R')$. Consequently, for every $D \in \text{Obj}\,\mathfrak{D}$, $h_{RD} \cong h_{RD'}$ and so by Yoneda, $RD \cong R'D$. Let $\gamma_D : RD \to R'D$ be the ismorophism with $h^C \gamma_D = (\beta\alpha)_{C,D}$, and let $f : D \to E$ be an arrow in $\mathfrak{D}$. Consider that the diagram below left commutes iff the diagram below right commutes, as h^C is fully faithful for every $C \in \text{Obj}C$, by Yoneda.

$$\begin{array}{ccc} h^C RD & \xrightarrow{(\beta\alpha)_{C,D}} & h^C R'D \\ {\scriptstyle h^C Rf}\downarrow & & \downarrow{\scriptstyle h^C R'f} \\ h^C RE & \xrightarrow[(\beta\alpha)_{C,E}]{} & h^C R'E \end{array} \qquad \begin{array}{ccc} RD & \xrightarrow{\gamma_D} & R'D \\ {\scriptstyle Rf}\downarrow & & \downarrow{\scriptstyle R'f} \\ RE & \xrightarrow[\gamma_E]{} & R'E \end{array}$$

However, the diagram on the left commutes by the naturality of $(\beta\alpha)$ and so $R \cong R'$ via γ, naturally. That left adjoints are unique follows by dualisation. ■

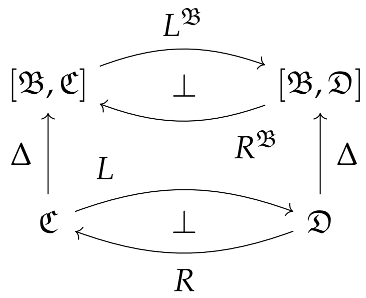
Prop. (A) 1.0.16. *If $L \dashv R : \mathfrak{B} \to \mathfrak{C}$ and $L' \dashv R' : \mathfrak{C} \to \mathfrak{D}$ then $L'L \dashv R'R : \mathfrak{B} \to \mathfrak{D}$.*

Proof. $\mathfrak{D}(L'LB, D) \cong \mathfrak{C}(LB, R'D) \cong \mathfrak{B}(B, R'RD)$. ■

Prop. (A) 1.0.17. *If $\mathfrak{C}$ has all limits of shape $\mathfrak{B}$, then $\Delta \dashv \lim : \mathfrak{C} \to [\mathfrak{B}, \mathfrak{C}]$.*

Proof. Limits are easily to seen to be a special case of universal arrows and so the statement is an immediate consequence of prop. (A) 1.0.14 (3). ■

Prop. (A) 1.0.18. *If $L \dashv R : \mathfrak{C} \to \mathfrak{D}$ then $L^{\mathfrak{B}} \dashv R^{\mathfrak{B}} : [\mathfrak{B}, \mathfrak{C}] \to [\mathfrak{B}, \mathfrak{D}]$ and the following diagram commutes, for any category $\mathfrak{B}$.*



Proof. Let $L \dashv R$ have unit and counit η, ϵ and define $\eta^{\mathfrak{B}} : \mathrm{id}_{[\mathfrak{B},\mathfrak{C}]} \to R^{\mathfrak{B}} L^{\mathfrak{B}}$ through components as ${\eta^{\mathfrak{B}}}_F = \eta F$, and similarly ${\varepsilon^{\mathfrak{B}}}_G = \varepsilon G$ for $F \in [\mathfrak{B}, \mathfrak{C}]$ and $G \in [\mathfrak{B}, \mathfrak{D}]$. We show that $\eta^{\mathfrak{B}}, \varepsilon^{\mathfrak{B}}$ are natural and that the triangle identities are obeyed.

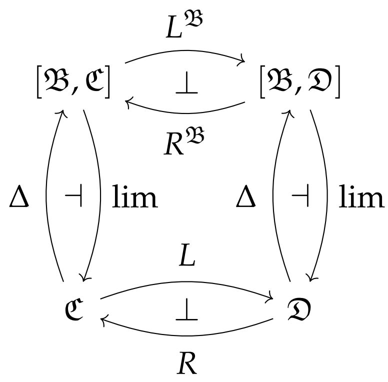
To see that $\eta^{\mathfrak{B}}$ is natural, let $F, F' \in [\mathfrak{B}, \mathfrak{C}]$ and $\tau : F \to F'$ natural between them. We wish to have ${\eta^{\mathfrak{B}}}_F R^{\mathfrak{B}} L^{\mathfrak{B}} \tau = \tau {\eta^{\mathfrak{B}}}_{F'}$. However, the left hand side is just $(\eta F)(RL\tau) = \tau \eta_{F'}$ by naturality of η. The proof for $\varepsilon^{\mathfrak{B}}$ is entirely similar.

That the triangle identities hold is equally trivial, as we are merely dealing with η and ε with components as the image of a functor. Specifically, we may expand $(\varepsilon^{\mathfrak{B}} L^{\mathfrak{B}})_F (L^{\mathfrak{B}} \eta^{\mathfrak{B}})_F = (\varepsilon LF)(L\eta F) = (\varepsilon L)(L\varepsilon)F = F$.

Finally, that the diagram commutes is also something of a triviality, in that we wish to show that $L^{\mathfrak{B}} \Delta = \Delta L$ and $R^{\mathfrak{B}} \Delta = \Delta R$. Of course, $L^{\mathfrak{B}} \Delta C = L \Delta C = \Delta LC$ for every $C \in \operatorname{Obj} \mathfrak{C}$, thereby concluding the proof. ■

Prop. (A) 1.0.19. *Right adjoints are continuous.*

Proof. Let $L \dashv R : \mathfrak{C} \to \mathfrak{D}$ and $\mathfrak{B}$ be a small category with $\mathfrak{C}$ and $\mathfrak{D}$ having all limits of shape $\mathfrak{B}$, and consider the following diagram.



In order to show the continuity of R, we must show that if $\alpha : \Delta \lim F \to F$ is the limiting cone for F then it must be that $R\alpha : R\Delta \lim F \to RF$ is the limiting cone for RF. Thus, let $\beta : \Delta C \to RF$ be a cone over RF and ${\Phi^{-1}}_{\Delta C,F} : [\mathfrak{B}, \mathfrak{C}](\Delta C, R^{\mathfrak{B}}F) \to [\mathfrak{B}, \mathfrak{D}](L^{\mathfrak{B}}\Delta C, F)$ be the binatural isomorphism arising from $L^{\mathfrak{B}} \dashv R^{\mathfrak{B}}$. As Φ^{-1} is an isomorphism, it is clear that ${\Phi^{-1}}_{\Delta C,F}\beta : \Delta LC \to F$ is a cone for F. As such, there exists a unique $u : LC \to \lim F$ such that ${\Phi^{-1}}_{\Delta C,F}\beta = \alpha\Delta u$.

If we write $\phi_{C,\lim F} : \mathfrak{D}(LC, \lim F) \to \mathfrak{C}(C, R \lim F)$ as the binatural isomorphism arising from $L \dashv R$, then we may note that $\phi_{C,\lim F}u : C \to R \lim F$ and further that, for $B \in \operatorname{Obj} \mathfrak{B}$

$$\begin{aligned} (R\alpha)_B \Delta(\phi_{C,\lim F} u) &= R\alpha_B \phi_{C,\lim F} u \\ &= \phi_{C,FB}(\alpha_B u) && \text{(naturality of } \phi_{C,-}) \\ &= \phi_{C,FB} \left({\Phi^{-1}}_{\Delta C,F}\beta\right)_B && ({\Phi^{-1}}_{\Delta C,F}\beta = \alpha\Delta u) \\ &= \phi_{C,FB}{\phi^{-1}}_{C,FB}\beta_B && (*) \\ &= \beta_B \end{aligned}$$

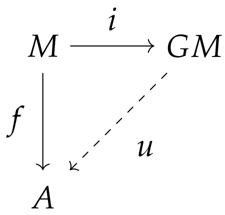
To see the equality marked $(*)$ we recall that ${\Phi^{-1}}_{\Delta C,F} = \varepsilon^{\mathfrak{B}} L^{\mathfrak{B}}$. With this in hand, to show that $R\alpha$ is a limiting cone has been reduced to the task of showing that $\phi_{C,FB}u$ is unique.

Suppose there was a $v : C \to R \lim F$ such that $R\alpha\Delta v = \beta$, then we check $\alpha_B {\phi^{-1}}_{C,\lim F} v = {\phi^{-1}}_{C,FB}\alpha_B v = \phi_{C,FB}\beta_B$ and so $\alpha\Delta{\phi^{-1}}_{C,\lim F} v = \Phi_{\Delta C,F}\beta$, forcing the equality ${\phi^{-1}}_{C,\lim F} v = u$ by universality of u for cones over F. Ergo, $R\alpha$ is a limiting cone for $R \lim F$.

Finally, we know that we may compose adjoints to find $L^{\mathfrak{B}}\Delta \dashv \lim R^{\mathfrak{B}}$, but we have $L^{\mathfrak{B}}\Delta = \Delta L$ and $\Delta L \dashv R \lim$ and so by prop. (A) 1.0.15 it must be the case that $\lim R^{\mathfrak{B}} \cong R \lim$ as functors $[\mathfrak{B}, \mathfrak{C}] \to \mathfrak{D}$. Thus, for any given functor $F : \mathfrak{B} \to \mathfrak{D}$, $\lim RF \cong R \lim F$. ∎

By way of completing what we wish to say about adjoints, we provide a specific case of adjoint functors as they arise "in nature".

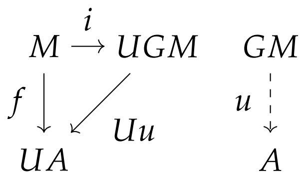
Def. (A) 1.0.20. The Grothendieck group GM of a commutative monoid M is the abelian group such that there exists a monoid homomorphism $i : A \to G$ which is universal with respect to the following property



where A is an abelian group, f is a monoid homomorphism, and u is a group homomorphism.

Remark (A) 1.0.21. There are several, equivalent ways to see that the Grothendieck group of a commutative monoid always exists, perhaps the most straightfoward of which is to define $G = M \times M/\sim$ where $(a, b) \sim (c, d) \iff (\exists k \in M)\, a + d + k = b + c + k$, and the group structure on the quotient as obvious. From this particular construction it is clear that M is cancellative iff $i : M \to G$ is injective.

Given the above remark, we are motivated to rewrite the definition in terms of the 'forgetful' functor U : Ab $\to$ CMon so as to restate the Grothendieck group in terms of more familiar language. Thus, wee see that the Grothendieck group map on objects G : CMon $\to$ Ab has the following universal property



and so, as we know, G extends to a functor and as a simple consequence,

Prop. (A) 1.0.22. $G \dashv U$: CMon $\to$ Ab

Proof. Prop. (A) 1.0.14 (3). ■

2. Regular categories

Although we managed to retrieve a great deal of algebraic notions of images and kernels in a more general manner, we did so in the setting of abelian categories.

In general it is undesirable to require the full force of abelianness (or even Ab enrichment), for example, to prove statements such as (but not limited to) unique factorisation of morphisms through their images. To this end, and for completion and to better flesh out the hierarchy presented, we achieve 'algebraic feeling' categories by introducing the following notion.

Def. (A) 2.0.23. Given an arrow $f : A \to B$ in $\mathfrak{C}$, the kernel pair of f is the pullback of f along itself, viz., $p_1, p_2 : P \rightrightarrows A$.

Remark (A) 2.0.24. In general categories, the kernel pair and the kernel are not isomorphic. In $\text{Set}_\bullet$, for example, the kernel of a morphism $f : (X, x) \to (Y, y)$ is the set $\{a \in X \mid f(a) = y\}$ whereas the kernel pair is the set $\{(a, a') \in X \times X \mid f(a) = f(a')\}$. The only "natural" map here is the diagonal inclusion of the former into the latter. In fact, this morphism exists in a general category, due to the universal property of the pullback.

Prop. (A) 2.0.25. *If the kernel pair (P, p_1, p_2) of $f : A \to B$ exists, then p_1 and p_2 are epimorphisms.*

Proof. Note that $f\,\text{id}_A = f\,\text{id}_A$ and so by the universal property of the pullback there is a unique morphism $u : A \to P$ such that $p_i u = \text{id}_A$. Thus u is a split monomorphism and p_i are split epimorphisms. ∎

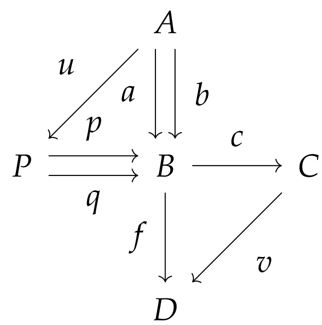
Prop. (A) 2.0.26. *The following conditions are equivalent for a morphism $f : A \to B$*

1. *f is a monomorphism*
2. *the kernel pair of f exists and is $(A, \text{id}_A, \text{id}_A)$*
3. *the kernel pair of f, (P, p_1, p_2) exists and has $p_1 = p_2$*

Proof. Assume that f is a monomorphism, and select a triple (C, c_1, c_2) with $fc_1 = fc_2$. We immediately have $c_1 = c_2$ and so may set $u = c_1 = c_2$ and uniqueness is evident. Then (2) obviously implies (3) and, assuming (3), given $a, b : D \rightrightarrows A$ with $fa = fb$ there is a unique arrow $u : D \to P$ with $b = up_i = a$. ∎

Prop. (A) 2.0.27. *If a coequaliser has a kernel pair, then it is the coequaliser of its kernel pair. If a kernel pair has a coequaliser, then it is the kernel pair of its coequaliser.*

Proof. With the diagram below, consider the following.



Suppose $c = \text{coeq}(a, b)$ and (P, p, q) is its kernel pair. By the pullback property, we have $u : A \to B$ with $pu = a$ and $qu = b$. Suppose further that $f : B \to D$ has $fp = fq$ then we have $fa = fpu = fqu = fb$ and so a unique arrow $v : C \to D$, by the coequaliser property, with $vc = f$ making $c = \text{coeq}(p, q)$ by universal property.

Suppose now that (P, p, q) is the kernel pair of f and $c = \text{coeq}(p, q)$. By the coequaliser property, we have $v : C \to D$ with $vc = f$. Suppose further that the parallel arrows $a, b : A \rightrightarrows B$ have $ca = cb$, then we have that $fa = vca = vcb = fb$ and so we must have a unique arrow $u : A \to P$, by the pullback property, with $a = pu$ and $b = qu$, making (P, p, q) the kernel pair of c by universal property. ■

There is a final, technical result that we exhibit before addressing the matter at heart of this section.

Lem. (A) 2.0.28 (Pasting lemma for pullbacks)**.** *In the following commutative diagram, where the right-hand square is a pullback, the left-hand square is a pullback iff the outer square is a pullback.*

$$\begin{array}{ccccc} A & \xrightarrow{a_B} & B & \xrightarrow{b_C} & C \\ {\scriptstyle a_D}\downarrow & & \downarrow{\scriptstyle b_E} & & \downarrow{\scriptstyle c} \\ D & \xrightarrow[d]{} & E & \xrightarrow[e]{} & F \end{array}$$

Proof. Suppose that the left-hand square is a pullback and that there is a G with arrows $f : G \to C$ and $g : G \to D$ such that $cf = edg$. Then in particular we can view G has having arrows $f : G \to C$ and $dg : G \to E$ onto the right-hand square such that $cf = e(dg)$, so that there exists a unique $u : G \to B$ such that $b_C u = f$ and $b_E u = dg$. Then, we may view G as having arrows $g : G \to D$ and $u : G \to B$ onto the left-hand square such that $dg = b_E u$ and so by the pullback property there exists a unique arrow $v : G \to A$ such that $a_B v = u$ and $a_D v = g$. Consequently, $b_C a_B v = b_C u = f$ and $a_D v = g$ with $cf = edg$ and so the outer square is a pullback.

Now suppose that the outer square is a pullback and that there is a G with arrows $f : G \to C$ and $g : G \to D$ such that $cf = edg$. In this case, there exists a unique arrow $v : G \to A$ such that $b_C a_B v = f$ and $a_D v = g$. As such, G may be viewed as having arrows $f : G \to C$ and $dg : G \to E$ such that $cf = e(dg)$. Then, by the universal property of the pullback there exists a unique arrow $u : G \to B$ such that $b_C u = f$ and $b_E u = dg$. Observe that $a_B v = u$ by the uniqueness of this arrow, as $b_C a_B v = f$ and $b_E a_B v = d a_D v = dg$ by commutativity and universality of v. As such, G may be viewed as having arrows $g : G \to D$ and $u : G \to B$ such that $b_E u = dg$, where there exists a unique arrow $v : G \to A$ such that $a_D v = g$ and $a_B v = u$. ■

With that established, we now define a category that straddles the gap between being algebraic in a structural way and demonstrating desirable properties for certain objects.

Def. (A) 2.0.29. A category is regular if the following hold

1. every arrow has a kernel pair
2. every kernel pair has a coequaliser
3. the pullback of a regular epimorphism along any morphism exists and is again a regular epimorphism

Conveniently, and the author assures the reader here that this is no accident, we have a wealth of 'good' examples of regular categories. As we can see, abelian categories are, in particular, regular.

Cor. (A) 2.0.30. *In a regular category, the pullback of a composite of regular epimorphisms is again a composite of regular epimorphisms.*

Proof. Let $a : A \to B$ and $b : B \to C$ be regular epimorphisms and $f : D \to C$ be an arbitrary arrow. Consider the pullbacks

$$\begin{array}{ccc} P & \xrightarrow{p_B} & B \\ {\scriptstyle p_D}\downarrow & & \downarrow{\scriptstyle b} \\ D & \xrightarrow[f]{} & C \end{array} \qquad\qquad \begin{array}{ccc} Q & \xrightarrow{q_B} & A \\ {\scriptstyle q_P}\downarrow & & \downarrow{\scriptstyle a} \\ P & \xrightarrow[p_B]{} & B \end{array}$$

where p_D and q_P are regular epimorphisms because the category is regular. It is apparent that we can paste these two diagrams together, and so by lem. (A) 2.0.28 the outer square must also be a pullback and so $q_P p_D : Q \to D$ is the composite of two regular epimorphisms. ■

Prop. (A) 2.0.31. *If $\mathfrak{C}$ is regular, and the kernel pair of $f : X \to Y$ is $p_0, p_1 : P \rightrightarrows X$, with $c : X \to C = \operatorname{coeq}(p_0, p_1)$, then the unique arrow $v : C \to Y$ that arises from the coequaliser via f is a monomorphism. That is, the following diagram commutes.*

$$\begin{array}{ccccc} P & \underset{p_1}{\overset{p_0}{\rightrightarrows}} & X & \xrightarrow{c} & C \\ & & {\scriptstyle f}\downarrow & \swarrow{\scriptstyle v} & \\ & & Y & & \end{array}$$

Proof. Suppose there were parallel arrows $g, h : A \rightrightarrows C$ such that $vg = vh$. We begin by taking the following pullback,

$$\begin{array}{ccc} B & \xrightarrow{q_0 \times q_1} & X \times X \\ {\scriptstyle a}\big\downarrow & & \big\downarrow{\scriptstyle c \times c} \\ A & \xrightarrow[g \times h]{} & C \times C \end{array}$$

Observe that $fq_0 = vcq_0 = vga = vha = vcq_1 = fq_1$ and so we may derive a unique arrow $u : B \to P$ due to the universal property of the kernel pair, such that $p_0 u = q_0$ and $p_1 u = q_1$. With this in hand, we may state that $ga = cq_0 = cp_0 u = cp_1 u = cq_1 = ha$, where the middle equality arises from the coequaliser nature of c. If it were the case that a was an epimorphism, then we would have $g = h$ and the proof would be completed.

In order to demonstrate this, we decompose $c \times c$. Note that $c \times c$ is the composite of $c \times \mathrm{id}_X : X \times X \to C \times X$ and $\mathrm{id}_C \times c : C \times X \to C \times C$. Moreover, both of these morphisms are themselves pullbacks,

$$\begin{array}{ccc} X \times X & \xrightarrow{\pi_0} & X \\ {\scriptstyle c \times \mathrm{id}_X}\big\downarrow & & \big\downarrow{\scriptstyle c} \\ C \times X & \xrightarrow[\pi_0]{} & C \end{array} \qquad \begin{array}{ccc} C \times X & \xrightarrow{\pi_1} & X \\ {\scriptstyle \mathrm{id}_C \times c}\big\downarrow & & \big\downarrow{\scriptstyle c} \\ C \times C & \xrightarrow[\pi_1]{} & C \end{array}$$

Then, because $\mathfrak{C}$ is regular, both $c \times \mathrm{id}_X$ and $\mathrm{id}_C \times c$ are regular epimorphisms. As such, cor. (A) 2.0.30 informs us that $a : B \to A$ is thus composite of two regular epimorphisms and so, in particular, an epimorphism, and the result follows. ■

Prop. (A) 2.0.32. *If $\mathfrak{C}$ is regular then every arrow $f : X \to Y$ with image has that $\mathrm{im}\, f \cong C = \mathrm{coeq}(p_0, p_1)$ where $p_0, p_1 : P \rightrightarrows X$ is the kernel pair.*

Proof. Consider the following commuting diagram which we recover through the definition of $\mathrm{im}\, f$ and prop. (A) 2.0.31

$$\begin{array}{ccccc} P & \overset{p_0}{\underset{p_1}{\rightrightarrows}} & X & \xrightarrow{c} & C \\ & & {\scriptstyle h}\big\downarrow & \overset{f}{\searrow} & \big\downarrow{\scriptstyle v} \\ & & \mathrm{im}\, f & \xrightarrow[g]{} & Y \end{array}$$

Observe that $ghp_0 = fp_0 = fp_1 = ghp_1$ and so $hp_0 = hp_1$ as g is a monomorphism. Thus we recover a unique monomorphism $v : C \to \mathrm{im}\, f$ by the coequaliser property. Moreover, as $v : C \to Y$ is a subobject through which f factors, there must be a unique monomorphism $w : \mathrm{im}\, f \to C$ by the image property. The result follows with little effort. ■

Given this, in a regular category when we speak of image we will in fact be referring to the coequaliser of the kernel pair, as an appropriate generalisation of image.

Prop. (A) 2.0.33. *If $\mathfrak{C}$ is regular then every $f : X \to Y$ can be factored uniquely through its image as $f = ie$ with $i : \operatorname{im} f \to Y$ a monomorphism and $e : X \to \operatorname{im} f$ a regular epimorphism.*

Proof. Let $p_0, p_1 : P \rightrightarrows X$ be the kernel pair of f, and let $e : X \to \operatorname{im} f$ be its coequaliser. It is evident that e is a regular epimorphism and, from prop. (A) 2.0.31, that the unique arrow $i : \operatorname{im} f \to Y$ is a monomorphism.

For uniqueness, suppose $f = ie = i'e'$ with $i' : I \to Y$ a monomorphism and $e' : X \to I$ a regular epimorphism as the coequaliser of $k, l : C \rightrightarrows X$. Note that $i'e'p_1 = fp_1 = fp_0 = i'e'p_0$ and so $e'p_0 = e'p_1$ as i' is a monomorphism. Thus, by the coequaliser property of e we have a unique arrow $a : \operatorname{im} f \to I$ such that $e' = ae$. Similarly, as e' is a coequaliser and $iek = fk = fl = iel$ we have a unique arrow $b : I \to \operatorname{im} f$ such that $e = be'$. Consequently, $e = be' = bae$ and so $ba = \mathrm{id}_{\operatorname{im} f}$. Similarly, $e' = ae = abe'$ and so $ab = \mathrm{id}_I$. All that remains to be done is to note that $i'e' = i'ae = f = ie$ so that $i = i'a$. ■

With the statement of this proof, the reader should be fully convinced that regular categories afford us one of the key luxuries of abelian categories without requiring nearly as much of the category in question. In particular then, it should come as no surprise that

Thm. (A) 2.0.34. *All abelian categories are regular.*

Proof. Unfortunately, the proof of this matter would lead us too far astray, but the reader is encouraged to consult [Bor94] for details. ■

Moreover, we may wonder just how much of our discussion of exactness may be recovered in the regular case. After all, we do not have any obvious way to speak of sums or differences here. It may come as something of a surprise then that we are able to define and prove the following.

Def. (A) 2.0.35. In a regular category, a diagram of the form

$$P \underset{q}{\overset{p}{\rightrightarrows}} A \xrightarrow{f} B$$

is said to be exact when (P, p, q) is the kernel pair of f and $f = \operatorname{coeq}(p, q)$.

Observe that, in the above, f is a regular epimorphism and so by the regularity of the category, p and q are regular epimorphisms. Of course, given that we know that abelian categories are regular, we may at this point be wondering whether exactness as defined above coincides with the standard definition in that context. To answer this question, and end off this section, we cite a result given in [Bor94] which shows that it is indeed the case.

Thm. (A) 2.0.36 (Borceux). *In an abelian category,*

$$P \underset{q}{\overset{p}{\rightrightarrows}} A \xrightarrow{f} B$$

is exact (in the above sense) iff the following is a short exact sequence

$$0 \longrightarrow P \xrightarrow{\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)} A \oplus A \xrightarrow{\left(\begin{smallmatrix} f & -f \end{smallmatrix}\right)} B \longrightarrow 0$$

Thus we have seen that it is possible, on the shoulders of weaker assumptions, to recover the unique epi-mono factorisation of morphisms and even provide a means (though perhaps less generally useful) of discussing exact morphisms.

Bibliography

[Ama98] S.-I. Amari. Natural gradient works efficiently in learning. *Neural Comput.*, 10(2):251–276, Feb. 1998.

[AN00] S-I. Amari and H. Nagaoka. *Methods of information geometry*, volume 191 of *Translations of mathematical monographs*. Am. Math. Soc., 2000.

[Bor94] F. Borceux. *Handbook of Categorical Algebra*, volume II. Cambridge University Press, 1994.

[Fre64] P. J. Freyd. *Abelian Categories*. Harper and Row, 1964.

[JS68] A. Joyal and R. Street. Braided monoidal categoreis. *Macquarie Math Reports*, (860081), 1968.

[Kel82] G. M. Kelly. *Basic Concepts of Enriched Category Theory*. Number 64 in Lecture Notes in Mathematics. Cambridge University Press, 1982.

[KL51] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22(1):79–86, 1951.

[McL90] C. McLarty. The uses and abuses of the history topos theory. *Brit. J. Phil. Sci.*, 31:351–375, 1990.

[ML97] S. Mac Lane. *Categories for the Working Mathematician*. Number 5 in Graduate Texts In Mathematics. Springer, 2nd edition, 1997.

[Nas56] J. Nash. The imbedding problem for riemannian manifolds. *Ann. Math.*, 63(1):20–63, January 1956.

[Rie11] E. Riehl. A leisurely introduction to simplicial sets. Personal website, `http://www.math.harvard.edu/~eriehl/ssets.pdf`, Aug. 2011.