TRENDS IN MATHEMATICS

Categorical Perspectives

Jürgen Koslowski
Austin Melton
Editors

Springer Basel AG

Trends in Mathematics

Series Titles

Anderson/Lapidus (eds). Progress in Inverse Spectral Geometry

Jain/Rizvi (eds). Advances in Ring Theory

Martino/Kantor/Lunardon/Pasini/Tamburini (eds). Groups and Geometries

Ross/Anderson/Litvinov/Singh/Sunder/Wildberger (eds). Harmonic Analysis and Hypergroups

Karatzas/Rajput/Taqqu (eds). Stochastic Processes and Related Topics: In Memory of Stamatis Cambanis 1943–1995

Komatsu/Kuranishi (eds). Analysis and Geometry in Several Complex Variables: Proceedings of the 40th Taniguchi Symposium

Birget/Margolis/Meakin/Sapir (eds). Algorithmic Problems in Groups and Semigroups

Ellingsrud/Fulton/Vistoli (eds). Recent Progress in Intersection Theory

Passi (ed). Algebra: Some Recent Advances

Eklof/Göbel (eds). Abelian Groups and Modules

Gyr/Kinzelbach/Tsinober (eds). Fundamental Problematic Issues in Turbulence

Bambah/Dumir/Hans-Gill (eds). Number Theory

Rebolledo (ed). Stochastic Analysis and Mathematical Physics: ANESTOC '98, Proceedings of the Third International Workshop

Hida/Karandikar/Kunita/Rajput/Watanabe/Xiong (eds). Stochastics in Finite and Infinite Dimensions: In Honor of Gopinath Kallianpur

Birkenmeier/Park/Park (eds). International Symposium on Ring Theory

Koslowski/Melton (eds). Categorical Perspectives

Trends in Mathematics

Trends in Mathematics is a book series devoted to focused collections of articles arising from conferences, workshops or series of lectures.

Topics in a volume may concentrate on a particular area of mathematics, or may encompass a broad range of related subject matter. The purpose of this series is both progressive and archival, a context in which to make current developments available rapidly to the community as well as to embed them in a recognizable and accessible way.

Volumes of TIMS must be of high scientific quality. Articles without proofs, or which do not contain significantly new results, are not appropriate. High quality survey papers, however, are welcome. Contributions must be submitted to peer review in a process that emulates the best journal procedures, and must be edited for correct use of language. As a rule, the language will be English, but selective exceptions may be made. Articles should conform to the highest standards of bibliographic reference and attribution.

The organizers or editors of each volume are expected to deliver manuscripts in a form that is essentially "ready for reproduction." It is preferable that papers be submitted in one of the various forms of TEX in order to achieve a uniform and readable appearance. Ideally, volumes should not exceed 350-400 pages in length.

Proposals to the Publisher are welcomed at either:
Birkhäuser Boston, 675 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.
math@birkhauser.com
or
Birkhäuser Verlag AG, PO Box 133, CH-4010 Basel, Switzerland
math@birkhauser.ch

Categorical Perspectives

Jürgen Koslowski
Austin Melton
Editors

Springer Science+Business Media, LLC

Jürgen Koslowski
Institut für Theoretische Informatik
TU Braunschweig
38023 Braunschweig
Germany

Austin Melton
Department of Mathematics
and Computer Science
Kent State University
Kent, OH 44242
U.S.A.

Library of Congress Cataloging-in-Publication Data

Categorical perspectives / Jürgen Koslowski, Austin Melton, editors.
p. cm.– (Trends in mathematics)
Includes bibliographical references.
ISBN 978-1-4612-7117-8 ISBN 978-1-4612-1370-3 (eBook)
DOI 10.1007/978-1-4612-1370-3

1. Categories (Mathematics) I. Koslowski, Jürgen, 1956- II. Melton, Austin. III. Series.

QA169.C35 2001
511.3–dc21

2001025482
CIP

AMS Subject Classifications: 03E25, 05C10, 05C76, 06A15, 18-01, 18-02, 18A02, 18A05, 18A10, 18A15, 18A20, 18A22, 18A30, 18A32, 18A35, 18A40, 18B05, 18D15, 54A05, 54A20, 54C35, 54D20, 54D30, 54E05, 54E15

Printed on acid-free paper

Originally published by Birkhäuser Boston in 2001
Softcover reprint of the hardcover 1st edition 2001

ISBN 978-1-4612-7117-8 SPIN 10761690

Typeset by the editors in LaTeX

9 8 7 6 5 4 3 2 1

Dedicated to George E. Strecker
on the occasion of his sixtieth birthday

Contents

Preface

This book had its beginnings in a conference held in honor of Professor George E. Strecker's 60th birthday. The conference was held in August 1998 on the campus of Kent State University.

During the planning for the conference, questions about a conference proceedings came up. It was not, however, until discussions with Birkhäuser started that the final format for this book began to take shape. Bernd Schroeder introduced us to Joseph Kung who helped us prepare materials for Birkhäuser and introduced us to Ann Kostant, Executive Editor for Mathematics and Physics at Birkhäuser. She and Tom Grasso, also of Birkhäuser, encouraged us to have expository and research chapters, and in discussions with prospective authors, the idea of including *teaching* chapters emerged. The end result is a delightful combination of teaching, expository, and research papers.

One of the teaching chapters, *10 Rules for Surviving as a Mathematician and Teacher* presents in a humorous and insightful manner suggestions for being a successful mathematician and a successful teacher. The other teaching chapters, written by Alois Zmrzlina and Y.T. Rhineghost, are tutorials in category theory; these tutorials are presented as dialogs between master teachers and their students.

Some of the chapters are mainly tutorial in nature, and some are mainly research. Of course, many are a combination of the two. Since a major emphasis of Professor Strecker's research has been category theory and its uses to explain and clarify other areas, mainly topology, it is natural that many chapters in this book use category theory in a similar manner – thus, the title *Categorical Perspectives.*

An introduction to category theory in general, and to this book in particular, is the chapter "Categories: A Free Tour" by Lutz Schröder. The chapter "Contributions and Importance of Professor George Strecker's Research" by Jürgen Koslowski is also a good introduction to this book and to the more specialized field of categorical topology.

We would like to extend a special *thank you* to Joyce Fuell, who was extremely helpful in formatting and preparing this book for publication.

Jürgen Koslowski
Institut für Theoretische Informatik
Technische Universitaät Braunschweig
Braunschweig, Germany
koslowj@iti.cs.tu-bs.de

Austin Melton
Department of Mathematics and Computer Science
Kent State University
Kent, OH 44242 (USA)
melton@mcs.kent.edu

Categorical Perspectives

Categories: A Free Tour

Lutz Schröder

ABSTRACT Category theory plays an important role as a unifying agent in a rapidly expanding universe of mathematics. In this paper, an introduction is given to the basic definitions of category theory, as well as to more advanced concepts such as adjointness, factorization systems and cartesian closedness.

Key words: Category, adjoint functor, cartesian closed category, factorization system.

AMS subject classifications (2000): 18-01, 18A05, 18A32, 18A40, 18D15.

In the past decades, the subject of mathematics has experienced an explosive increase both in diversity and in the sheer amount of published material. (E.g., the Mathematical Reviews volume of 1950 features 766 pages of reviews, compared to a total of 4550 pages in the six volumes for *the first half* of 2000.) It has thus become inevitable that this growth, taking place in numerous and increasingly disconnected branches, be complemented by some form of unifying theory. There have been attempts at such unifications in the past, such as Birkhoff-style universal algebra or the encyclopedic work of Bourbaki. However, the most successful and universal approach so far is certainly the theory of categories as initiated by Eilenberg and Mac Lane in 1945 [14].

Category theory is based on the observation that mathematical structures practically always come with a notion of morphism between these structures, and that these morphisms deserve at least as much attention as the structures themselves (this insight of category theory might well be the only one that has actually managed to pervade most of mathematics); e.g., one has homomorphisms between algebraic structures, continuous maps between topological structures, order-preserving or (!) continuous maps between order-theoretic structures. Moreover, two morphisms of the same type (with suitable domain and codomain, respectively) can usually be composed to yield another morphism. (For a generalization of the theory to cases where this property fails see [34, 35, 36].)

This — objects, morphisms, and composition — is more or less all there is to the definition of category; the only remaining step is that of abstracting from actual sets and maps towards mere arrows between nodes of an unspecified nature, subject to a binary algebraic operation just as, say, the elements of a group (which have in the course of time undergone a similar, if rather slower, abstraction process). The result is itself just a type of mathematical structure, incorporating both algebraic and order-theoretic traits.

From these extremely simple beginnings, a rich, diversified and, contrary to popular opinion, non-trivial theory (which, in turn, is on the verge of being in need of unification ...) has developed. It encompasses areas of application to existing and developing mathematics, e.g., categorical algebra and topology, as well as the abstract treatment of categories as objects of study worthy of interest for their own sake (including applications of categorical methods to category theory itself!).

The promised tour of category theory, which shall begin presently, will start off with an introduction to basic concepts such as category, functor, and natural transformation (Sections 1–3). We will then move on to adjoint functors (Section 4), which provide the proper axiomatization of free structures and universal properties as already present in Bourbaki [4]. This paves the way for a discussion of cartesian closed categories, which deal with an abstract notion of function space, in Section 6. Factorization structures, a rather more abstract tool which, among other things, serves to provide a sufficiently flexible notion of substructure, are treated in Section 5. The tour ends in Section 7 with a view of more topics ahead.

1 Categories

As indicated in the introduction, a category consists of objects (e.g., sets), morphisms between these objects (e.g., maps) and an operator that composes 'adjacent' morphisms (e.g., composition of maps). More formally:

Definition 1.1 A *category* is a quadruple $\mathbf{A} = (\mathcal{O}, hom, id, \circ)$, where

- $\mathcal{O} =: \mathrm{Ob}\,\mathbf{A}$ is a class, the members of which are called *objects*;
- *hom* assigns to each pair (A, B) of objects a set $hom(A, B)$ (or $hom_{\mathbf{A}}(A, B)$), the members of which are called the *morphisms from A to B* and denoted in the form $f : A \to B$;
- *id* assigns to each object A a morphism id_A from A to A, called the *identity on A*; and
- $\circ$ is an operator, called *composition*, which assigns to each pair of

morphisms $f : A \to B$ and $g : B \to C$ a *composite* $g \circ f : A \to C$:

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "g \circ f"'] & B \arrow[d, "g"] \\ & C \end{tikzcd} .$$

$g \circ f$ is often abbreviated as gf.

These data are subject to the following axioms:

- For each $f : A \to B$, $id_B f = f = f id_A$.
- Composition is *associative*, i.e., for $f : A \to B$, $g : B \to C$, and $h : C \to D$,

$$h(gf) = (hg)f.$$

- The sets $hom(A, B)$ are pairwise disjoint.

In certain special cases, this definition collapses into well-known concepts: categories with exactly one object are essentially monoids (the morphisms from the object to itself are the elements of a monoid with composition as the monoid operation and the identity on the single object as the unit), and categories where the sets $hom(A, B)$ have at most one element are essentially preordered classes (the class of objects is preordered by taking $A \leq B$ iff $hom(A, B) \neq \emptyset$). It does at times help to view categories as generalized forms of these examples, i.e., as monoids with more than one unit or as preorders in which an element can be smaller than another in more than one way. There is, in fact, a whole theory which deals with 'groups with more than one unit', i.e., with categories where each morphism is invertible in the obvious sense; such structures are called *groupoids* [5, 6]. (A typical example is the category which has the points of a given topological space as objects and homotopy classes of paths as morphisms.)

The main focus of category theory is, of course, set on 'real' categories such as the following:

Example 1.2 As indicated in the first paragraph above, one has a category **Set** consisting of all sets as objects and maps between sets as morphisms. Identities are identity maps, and composition is just composition of maps.

Example 1.3 Taking all groups as objects and homomorphisms between groups as morphisms produces a category **Grp**. Similarly, topological spaces and continuous maps form a category **Top**.

We have already seen examples of categories where the morphisms are not maps; here are some more complex examples of this kind:

Example 1.4 Morphisms may be equivalence classes of maps; e.g., topological spaces and homotopy classes of continuous maps form a category **hTop**.

Example 1.5 Morphisms may be tuples of maps; a simple example is the category **Set** $\times$ **Set**, which has pairs of sets (X, Y) as objects and pairs of maps $(f : X_1 \to X_2, g : Y_1 \to Y_2)$ as morphisms from (X_1, Y_1) to (X_2, Y_2). More generally, given two categories **A** and **B**, one has a *product category* $\mathbf{A} \times \mathbf{B}$ formed in the same way.

Example 1.6 Morphisms may be 'maps going the wrong way'; e.g., taking sets as objects and maps $f : Y \to X$ as morphisms *from* X *to* Y produces a category $\mathbf{Set}^{\mathbf{op}}$. More generally, given a category **A**, one obtains a category $\mathbf{A}^{op}$ by reversing the direction of all morphisms (and the order of composition). $\mathbf{A}^{op}$ is called the *dual category* of **A**. By virtue of this construction, every concept and every theorem in category theorem comes with a second version (called the *dual* definition or theorem, respectively), where all arrows are reversed. Dual concepts are most often named by just adding the prefix 'co-'.

Example 1.7 Morphisms may be partial maps or relations; e.g., sets and relations (with the usual composition of relations) form a category $\mathbf{Set_r}$.

Example 1.8 Morphisms may be something altogether different; e.g., given any directed graph, one can construct a category by taking its nodes as the objects, all paths between nodes as morphisms and concatenation of paths as composition.

Example 1.9 Morphisms may be completely absent; i.e., any class $\mathcal{C}$ can be made into a category which has the elements of $\mathcal{C}$ as objects and no morphisms other than the identities. Such categories are called *discrete.* (Of course, this is a special case of the interpretation of preordered classes as categories.) In particular, there is an empty category $\emptyset$.

We can now go on to define concepts in the language of categories and observe how they subsume well-known phenomena in concrete cases:

Definition 1.10 A morphism $f : A \to B$ in a category is called a *monomorphism* if, whenever $fg = fh$ for morphisms $g, h : C \to A$, then $g = h$. The dual notion (cf. Example 1.6) is called *epimorphism.* A monomorphism $f : A \to B$ is called *extremal* if, whenever $f = ge$ for some morphism $g : C \to B$ and some epimorphism $e : A \to C$, then e is an *isomorphism*, i.e., there exists $h : C \to A$ such that $eh = id_C$ and $he = id_A$ (it is easily seen that such an h, called the *inverse* of e, is unique).

Example 1.11 Morphisms in **Set**, **Grp** and **Top** are monomorphisms iff their underlying map is injective. A relation $\rho : X \to Y$ is a monomorphism in $\mathbf{Set_r}$ iff, for each $x \in X$, there exists $y \in Y$ such that x is unique with the property $x\rho y$.

Morphisms in **Set**, **Grp** and **Top** are epimorphisms iff their underlying map is surjective. In the category **Haus** of Hausdorff spaces and continuous maps, the epimorphisms are precisely the dense maps.

The extremal monomorphisms in **Grp** are, again, the injective homomorphisms. By way of contrast, the extremal monomorphisms in **Top** are essentially the subspace embeddings. In the category **Mon** of monoids and monoid homomorphisms, monomorphisms may also fail to be extremal: e.g., the embedding $\mathbb{Z} \hookrightarrow \mathbb{Q}$ (as multiplicative monoids) is itself an epimorphism.

While the appropriate categorical notions of subobject and quotient may to a certain extent vary from case to case, as the above example shows, one has a universally applicable notion of product:

Definition 1.12 Let $(A_i)_I$ be a (finite) family of objects in a category. A (finite) *product* of $(A_i)_I$ is an object A together with a family of projections $\pi_i : A \to A_i$, $i \in I$, such that, for each family $(f_i)_I$ of morphisms $f_i : B \to A_i$, there exists a unique morphism $f : B \to A$ such that $\pi_i f = f_i$ for each i:

$$\begin{tikzcd} A \arrow[r, "\pi_i"] & A_i \\ B \arrow[u, dotted, "f"] \arrow[ur, "f_i"'] & \end{tikzcd} .$$

A category **A** *has (finite) products* if there exists a product for each (finite) set-indexed family of objects in **A**.

It is an easy exercise to check that products are unique 'up to isomorphism'. (What, precisely, does this mean?) Products of pairs (A, B) are denoted in the form $A \times B$. What is a product of an empty family?

Example 1.13 Products in **Set** are cartesian products (with the usual projections), while products in $\mathbf{Set_r}$ are disjoint sums (with the inverse relations of the injections into the sum as projections). Products in **Top** are cartesian products with the well-known product topology. Products in **Grp** and other categories of algebraic nature are direct products, i.e., cartesian products with the operations defined component-wise.

The dual notion is that of a *coproduct*; the associated morphisms from the component objects *into* the coproduct are called *injections.* Coproducts in **Set** (and in $\mathbf{Set_r}$) are disjoint unions (together with the injections into the union). Coproducts in the category k-**Vec** of k-vector spaces and linear maps, where k is a field, are direct sums (so that the product and the coproduct of finitely many objects have the same object part). Coproducts in **Grp** are free products.

As in all mathematical theories, 'substructures' play a prominent role in category theory. There are at least two intuitively appealing notions of 'substructure of a category': one might either take a subclass of the object class and let it inherit 'everything else' from its parent, or, more generally, take a subclass of the morphism class with the appropriate closure properties. Both concepts are of elementary importance:

Definition 1.14 Let $\mathbf{A}$ be a category. A category $\mathbf{B}$ is a *subcategory* of $\mathbf{A}$ if $\mathrm{Ob}\,\mathbf{B}$ is a subclass of $\mathrm{Ob}\,\mathbf{A}$, $hom_{\mathbf{B}}(A,B) \subset hom_{\mathbf{A}}(A,B)$ for each pair (A,B) of objects in $\mathbf{B}$, and identities and composition in $\mathbf{B}$ are the same as in $\mathbf{A}$. $\mathbf{B}$ is called a *full subcategory* of $\mathbf{A}$ if, additionally, $hom_{\mathbf{B}}(A,B) = hom_{\mathbf{A}}(A,B)$ for each pair (A,B) of objects in $\mathbf{B}$.

A full subcategory $\mathbf{B}$ is called *isomorphism-closed* in $\mathbf{A}$ if, whenever $f : A \to B$ is an isomorphism in $\mathbf{A}$ and B is a $\mathbf{B}$-object, then A is a $\mathbf{B}$-object, and *closed under products* if, whenever A is the product of a family of $\mathbf{B}$-objects in $\mathbf{A}$, then A is a $\mathbf{B}$-object.

Example 1.15 **Haus** (cf. Example 1.11) is a full subcategory of **Top**. **Set** is a (non-full) subcategory of $\mathbf{Set_r}$. The category **Sgr** of semigroups and semigroup homomorphisms contains **Mon** (cf. Example 1.11) as a non-full subcategory (since semigroup homomorphisms need not preserve units of monoids) and **Grp** as a full subcategory (since preservation of the unit is automatic for groups).

There is a natural interest in subcategories with certain 'good' properties, the foremost of which is the following (or its dual):

Definition 1.16 Let $\mathbf{B}$ be a subcategory of a category $\mathbf{A}$. A $\mathbf{B}$-*reflection* of an object A in $\mathbf{A}$ is a morphism $r_A : A \to B_A$ into an object B_A in $\mathbf{B}$ such that, for each morphism $f : A \to B$ into an object B in $\mathbf{B}$, there exists a unique morphism $\bar{f} : B_A \to B$ in $\mathbf{B}$ such that $\bar{f} r_A = f$:

$$\begin{tikzcd} A \arrow[r, "r_A"] \arrow[dr, "f"'] & B_A \arrow[d, dotted, "\bar{f}"] \\ & B. \end{tikzcd}$$

$\mathbf{B}$ is called *reflective* if each $A \in \mathrm{Ob}\,\mathbf{A}$ has a $\mathbf{B}$-reflection.

(The reader is not mistaken in discerning a certain pattern in the definitions. In particular, the proof that reflective arrows are 'unique up to isomorphism' is analogous to the case of products.)

Example 1.17 The full subcategory **Ab** of **Grp** formed by the abelian groups and their homomorphisms is reflective in **Grp**; the reflection of a group is obtained by factoring out its commutator subgroup.

The category **HComp** of compact Hausdorff spaces is a full reflective subcategory of **Top**; the associated reflections are the Čech–Stone-compactifications.

The dual notion is that of a *coreflective* subcategory. E.g., **Set** is a non-full coreflective subcategory of $\mathbf{Set_r}$: given a set X (in $\mathbf{Set_r}$), its **Set**-coreflection is $\ni$: $\mathfrak{P}X \to X$, where $\mathfrak{P}X$ denotes the powerset of X.

The terminology collected so far can already be used to formulate and prove moderately deep but very general statements such as the following:

Theorem 1.18 *Isomorphism-closed full reflective subcategories are closed under products.*

In order to give the reader a taste of the straightforward simplicity of categorical arguments, we will, just this once, include a proof:

Proof. Let $\mathbf{B}$ be an isomorphism-closed full reflective subcategory of a category $\mathbf{A}$, let $(B_i)_I$ be a family of objects in $\mathbf{B}$, and let $(\pi_i : A \to B_i)_I$ be a product in $\mathbf{A}$. We have to show that A is a $\mathbf{B}$-object. To this end, let $r : A \to B_A$ be a $\mathbf{B}$-reflection of A. Then there exists, for each $i \in I$, a morphism $p_i : B_A \to B_i$ such that $p_i r = \pi_i$. By the product property, there exists $p : B_A \to A$ such that $\pi_i p = p_i$ for each i:

$$\begin{tikzcd} A \arrow[r, "\pi_i"] \arrow[d, shift right, "r"'] & B_i \\ B_A \arrow[u, shift right, "p"'] \arrow[ur, "p_i"'] & . \end{tikzcd}$$

Now $\pi_i(pr) = p_i r = \pi_i$ for each i, so that $pr = id_A$ by the uniqueness requirement in the definition of product. Hence $(rp)r = r$, which implies $rp = id$ by the uniqueness requirement in the definition of reflection. Thus p is an isomorphism, and hence A is a $\mathbf{B}$-object. ■

2 Functors

Now that the point that mathematical objects should always be considered in close connection with the morphisms between them has been sufficiently stressed, it comes as no surprise that morphisms *between categories* are an important concept. Such morphisms are called functors; they are, roughly speaking, maps between the involved categories that preserve identities and composition. More formally:

Definition 2.1 Let $\mathbf{A}$ and $\mathbf{B}$ be categories. A *functor* F from $\mathbf{A}$ to $\mathbf{B}$ consists of a map $F : \mathrm{Ob}\,\mathbf{A} \to \mathrm{Ob}\,\mathbf{B}$ and a family of maps $F : hom(A, B) \to hom(FA, FB)$, where A and B range over $\mathrm{Ob}\,\mathbf{A}$, such that

$$Fid_A = id_{FA}$$

for each object A and

$$FgFf = F(gf)$$

for each pair of morphisms $f : A \to B$ and $g : B \to C$ in $\mathbf{A}$.

Of course, one has an *identity functor* $id_{\mathbf{A}}$ on each category $\mathbf{A}$ (which acts as the identity on objects as well as on morphisms), and the composite of two functors in the obvious sense is again a functor. Thus, categories and functors almost form a category, except for the fact that the collection of all categories (as well as, in general, the collection of all functors between two categories) is 'too large', i.e., fails to form a class. *Small* categories, i.e., categories $\mathbf{A}$ such that $\mathrm{Ob}\,\mathbf{A}$ (and, hence, $\mathrm{Mor}\,\mathbf{A}$) is a set, and functors do form a category, which is denoted by **Cat**.

Example 2.2 Categories that consist of 'structured sets' as objects and 'structure preserving maps' as morphisms admit a functor to **Set** which maps a structured set to the underlying set and a morphism to its underlying map. Such functors are appropriately called *forgetful functors*; e.g., one has forgetful functors $\mathbf{Top} \to \mathbf{Set}$ and $\mathbf{Grp} \to \mathbf{Set}$. Despite their apparent triviality, properties of forgetful functors play an important role in many areas of category theory such as categorical algebra or, in particular, categorical topology.

Example 2.3 If $\mathbf{B}$ is a subcategory of a category $\mathbf{A}$, then the embedding $\mathbf{B} \hookrightarrow \mathbf{A}$ is a functor.

Example 2.4 If $\mathbf{I}$ is a class, considered as a discrete category (Example 1.9), then a functor from $\mathbf{I}$ to a category $\mathbf{A}$ is just an $\mathbf{I}$-indexed family of $\mathbf{A}$-objects.

Example 2.5 Given categories $\mathbf{A}$ and $\mathbf{B}$, one has obvious *projection functors* from the product category $\mathbf{A} \times \mathbf{B}$ (cf. Example 1.5) to $\mathbf{A}$ and $\mathbf{B}$.

Example 2.6 Let $\mathbf{A}$ and $\mathbf{B}$ be categories, and let B be an object in $\mathbf{B}$. Then there is a *constant functor* $\Delta B : \mathbf{A} \to \mathbf{B}$ which maps all objects to B and all morphisms to id_B.

Example 2.7 Given a category $\mathbf{A}$ and an object A in $\mathbf{A}$, one has a functor $hom(A, \_) : \mathbf{A} \to \mathbf{Set}$ which maps an object B to $hom(A, B)$ and a morphism $f : B \to C$ to the map that sends $g : A \to B$ to $fg : A \to C$. Dually, there is a functor $hom_{\mathbf{A}}(\_, A) : \mathbf{A}^{op} \to \mathbf{Set}$. These functors are unified in a functor $hom(\_,\_) : \mathbf{A}^{op} \times \mathbf{A} \to \mathbf{Set}$ which maps a pair (A, B) of objects to $hom(A, B)$ and a pair (f, g) of morphisms to the map $h \mapsto ghf$.

Example 2.8 Given a functor $F : \mathbf{A} \to \mathbf{B}$, one has a functor $F^{op} : \mathbf{A}^{op} \to \mathbf{B}^{op}$ which acts precisely as F.

Example 2.9 The assignment that maps each set to its power set can be extended to a functor in (at least) two ways: on the one hand, one has a *covariant power set functor* $\mathbf{Set} \to \mathbf{Set}$ which sends a map $f : X \to Y$ to the function $\mathfrak{P}X \to \mathfrak{P}Y$ which maps $A \subset X$ to $f[A]$, and on the other hand, there is a *contravariant power set functor* $\mathbf{Set}^{op} \to \mathbf{Set}$ which assigns to a map $f : Y \to X$ (i.e., a morphism $X \to Y$ in $\mathbf{Set}^{op}$) the map $\mathfrak{P}X \to \mathfrak{P}Y$ which maps $A \subset X$ to $f^{-1}[A]$.

Example 2.10 Assigning to a topological space its first homology group (and to a continuous map the associated homomorphism of homology groups) defines a functor $F : \mathbf{Top} \to \mathbf{Grp}$. The well-known homotopy invariance of homology means precisely that F factors through the functor $\mathbf{Top} \to \mathbf{hTop}$ (Example 1.4) which maps each continuous map to its homotopy class. The fact that F (as all functors do) *preserves isomorphisms* (i.e., if f is an isomorphism, then so is Ff) is the basis for the paradigmatic argument of algebraic topology, namely that two spaces with non-isomorphic homology groups cannot be homeomorphic (or even homotopy-equivalent).

Example 2.11 One has a functor $\mathbf{Set} \to \mathbf{Grp}$ which assigns to each set X the free group over X and to each map the associated 'renaming of variables'. Similarly (!), it can be seen that the construction of the abelian reflection of a group (cf. Example 1.17) extends to a functor $\mathbf{Grp} \to \mathbf{Ab}$. A general explanation for these phenomena is given in Section 4.

Unlike in most other mathematical theories, injectivity and, in particular, surjectivity of functors are of lesser interest compared to the following related notions:

Definition 2.12 A functor $F : \mathbf{A} \to \mathbf{B}$ is called *faithful* if $Ff = Fg$ implies $f = g$ for $f, g : A \to B$ in $\mathbf{A}$, and *full* if, whenever $g : FA \to FB$ is a morphism in $\mathbf{B}$, there exists a morphism $\bar{g} : A \to B$ such that $F\bar{g} = g$.

E.g., besides identity and embedding functors, the forgetful functors from Example 2.2, the power set functors, and the free group functor (Example 2.11) are faithful. Embeddings of full subcategories are, of course, full. An example of a functor that is full but not faithful is the 'quotient functor' $\mathbf{Top} \to \mathbf{hTop}$ from Example 2.10.

Similarly, actual *isomorphisms* between categories, i.e., functors that are bijective on morphisms (and, hence, on objects), are less ubiquitous than the (weaker) so-called *equivalences*, i.e., full and faithful functors $F : \mathbf{A} \to \mathbf{B}$ that are *isomorphism-dense* in the sense that each object B in $\mathbf{B}$ is isomorphic to some FA, $A \in \operatorname{Ob} \mathbf{A}$. Equivalent categories are regarded as being 'the same for the purposes of category theory'.

Example 2.13 Let **FinSet** denote the category of finite sets and maps, and let **Neu** denote the full subcategory of **FinSet** spanned by the von Neumann natural numbers. Then the inclusion $\mathbf{Neu} \hookrightarrow \mathbf{FinSet}$ is an equivalence.

Example 2.14 Let $\mathbf{Set_p}$ denote the category of sets and partial maps, and let **pSet** denote the category of pointed sets and point-preserving maps. The functor $\mathbf{pSet} \to \mathbf{Set_p}$ that sends a pointed set (X, p) (where p is the distinguished element) to the set $X - \{p\}$ and a point-preserving map $f : (X, p) \to (Y, q)$ to the partial function that is defined for $x \in X - \{p\}$ (and, then, acts as f) iff $f(x) \neq q$ is an equivalence (but not an isomorphism).

Further properties of functors that are typically of interest include *preservation* and *reflection* of properties of objects, morphisms and the like. One example is the preservation of isomorphisms mentioned in Example 2.10. As another example, it is easily seen that the forgetful functor $\mathbf{Grp} \to \mathbf{Set}$ *preserves products* in the obvious sense, but does not preserve coproducts. Reflection of a property means in general that an object, morphism etc. has the said property whenever its image does. E.g., the forgetful functor $\mathbf{Grp} \to \mathbf{Set}$, unlike the forgetful functor $\mathbf{Top} \to \mathbf{Set}$, reflects isomorphisms. Moreover, it is easily shown that faithful functors reflect monomorphisms. Equivalences preserve and reflect all 'categorical properties'.

3 Natural Transformations

Now that we have introduced functors as morphisms between categories, we can push the game one step further by considering morphisms *between functors*, more precisely between functors with the same domain and codomain. Such morphisms are called natural transformations. Arguably, natural transformations are what category theory really is about; at any rate, categories and functors were originally introduced by Eilenberg and MacLane [14] in order to provide a proper definition of naturality.

Definition 3.1 Let $F, G : \mathbf{A} \to \mathbf{B}$ be functors. A *natural transformation* $\mu : F \to G$ is a family of $\mathbf{B}$-morphisms $\mu_A : FA \to GA$, where A ranges over $\mathrm{Ob}\,\mathbf{A}$, such that, for each morphism $f : A \to B$ in $\mathbf{A}$, $\mu_B Ff = Gf\mu_A$:

$$\begin{tikzcd} FA \arrow[r, "\mu_A"] \arrow[d, "Ff"'] & GA \arrow[d, "Gf"] \\ FB \arrow[r, "\mu_B"'] & GB \end{tikzcd} \quad .$$

It is easily verified that the componentwise composition of two natural transformations is again a natural transformation. Moreover, families of identities id_{FA} are natural transformations $F \to F$, so that we have a *functor category* $[\mathbf{A}, \mathbf{B}]$ with functors $\mathbf{A} \to \mathbf{B}$ as objects and natural transformations as morphisms. In the general case, $[\mathbf{A}, \mathbf{B}]$ is not really a category, since its objects may fail to form a class; however, in the important case that $\mathbf{A}$ is small, $[\mathbf{A}, \mathbf{B}]$ is indeed a category.

For later reference, we note that, given functors $F, G : \mathbf{A} \to \mathbf{B}$, $H : \mathbf{C} \to \mathbf{A}$, and $K : \mathbf{B} \to \mathbf{D}$, a natural transformation $\mu : F \to G$ gives rise to natural transformations $\mu_H : FH \to GH$ and $K\mu : KF \to KH$ with components μ_{HC} and $K\mu_A$, respectively.

Example 3.2 The classical example of a natural transformation is the embedding of a vector space into its double dual: denoting, given a k-vector space V, its dual, i.e., the space of all linear maps $V \to k$, by V^*, we have a functor $\_^* : k\text{-}\mathbf{Vec} \to k\text{-}\mathbf{Vec}^{op}$ (Example 1.13. How does $\_^*$ act on morphisms?). Hence we have a functor $\_^{**} : k\text{-}\mathbf{Vec} \to k\text{-}\mathbf{Vec}$ which sends a vector space to its double dual, and the maps

$$\begin{array}{rcl} \phi_V : V & \to & V^{**}, \\ v & \mapsto & (f \mapsto f(v)) \end{array}$$

form a natural transformation $id_{k\text{-}\mathbf{Vec}} \to \_^{**}$.

Example 3.3 Given categories $\mathbf{A}$ and $\mathbf{B}$ and a morphism $f : A \to B$ in $\mathbf{B}$, there is a *constant natural transformation* $\Delta f : \Delta A \to \Delta B$ between constant functors (cf. Example 2.6), where $(\Delta f)_C = f$ for each $C \in \operatorname{Ob} \mathbf{A}$. Thus we have a *diagonal functor* $\Delta : \mathbf{B} \to [\mathbf{A}, \mathbf{B}]$.

Example 3.4 If $\mathbf{I}$ is a class, considered as a discrete category, then a natural transformation $F \to G$, where $F, G : \mathbf{I} \to \mathbf{A}$ are functors, is just an $\mathbf{I}$-indexed family of $\mathbf{A}$-morphisms. Thus, the functor category $[\mathbf{I}, \mathbf{A}]$ is essentially the cartesian product of an $\mathbf{I}$-indexed family of copies of $\mathbf{A}$ with the componentwise operations. (This is a product in the 'category' of all categories.)

Example 3.5 Denote the free group functor from Example 2.11 by F and the forgetful functor $\mathbf{Grp} \to \mathbf{Set}$ by U. Then the insertions of generators $X \hookrightarrow UFX$ form a natural transformation $id_{\mathbf{Set}} \to UF$. Similarly, the quotient maps w.r.t. the commutator subgroups form a natural transformation $id_{\mathbf{Grp}} \to ER$, where R denotes the 'abelianization functor' $\mathbf{Grp} \to \mathbf{Ab}$ from Example 2.11 and E denotes the inclusion $\mathbf{Ab} \hookrightarrow \mathbf{Grp}$.

Example 3.6 Take U as in the last example, and denote the functor that assigns the cartesian product $UG \times UG$ to each group G (and acts in the obvious way on morphisms) by U^2. The multiplication maps $UG \times UG \to UG$, $(a, b) \mapsto ab$, form a natural transformation $U^2 \to U$. This observation is central to the idea of algebraic theories [29, 30].

All categorical notions defined so far can, of course, be applied to functor categories. E.g., it is easily verified that isomorphisms in functor categories are precisely the *natural isomorphisms*, i.e., natural transformations the components of which are isomorphisms. (It has to be checked that the inverses of the components again form a natural transformation.)

Example 3.7 Let $Q : \mathbf{Set} \to \mathbf{Set}$ denote the *squaring functor* $X \mapsto X \times X$. Then one has a natural isomorphism $\eta : hom(2, \_) \to Q$ (Example 2.7), where 2 denotes the set $\{0, 1\}$; η_X maps $f : 2 \to X$ to the pair $(f(0), f(1))$.

Example 3.8 The natural transformation from the identity to the double dual functor on vector spaces (Example 3.2) becomes a natural isomorphism when restricted to the full subcategory of finite dimensional vector spaces. Thus we have now assigned a precise meaning to the well-known informal statement that, unlike in the case of the simple dual, a finite dimensional vector space is 'canonically' isomorphic to its double dual.

Example 3.9 The forgetful functor $\mathbf{Top} \to \mathbf{Set}$ is naturally isomorphic to $hom(1, \_)$, where 1 denotes the singleton topological space. Similarly (!), the forgetful functor $\mathbf{Grp} \to \mathbf{Set}$ is naturally isomorphic to $hom(\mathbb{Z}, \_)$, where $\mathbb{Z}$ denotes the additive group of integers. This is expressed by saying that these functors are *represented* by the objects 1 and $\mathbb{Z}$, respectively.

One of the central features of category theory is that functors not only form a category (analogous statements hold in many algebraic categories; e.g., the linear maps between two vector spaces form a vector space), but this category has a particularly interesting property (the formulation of which should, by now, ring a bell):

Theorem 3.10 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be categories. Then one has a functor*

$$ev : [\mathbf{A}, \mathbf{B}] \times \mathbf{A} \to \mathbf{B},$$

given by $ev(F, f) = Ff$ *for each functor* $F : \mathbf{A} \to \mathbf{B}$ *and each* $f \in \mathrm{Mor}\,\mathbf{A}$, *and* $ev(\mu, f) = Gf\mu_A = \mu_B Ff$ *for each* $\mu : F \to G$ *in* $[\mathbf{A}, \mathbf{B}]$. *Moreover, for each functor* $H : \mathbf{C} \times \mathbf{A} \to \mathbf{B}$, *where* $\mathbf{C}$ *is a category, there exists a unique functor* $\widehat{F} : \mathbf{C} \to [\mathbf{A}, \mathbf{B}]$ *such that* $ev(\widehat{F} \times id_{\mathbf{A}}) = F$*:*

$$\begin{tikzcd}
{[\mathbf{A}, \mathbf{B}] \times \mathbf{A}} \arrow[r, "ev"] & \mathbf{B} \\
\mathbf{C} \times \mathbf{A} \arrow[u, "\widehat{F} \times id_{\mathbf{A}}"] \arrow[ur, "F"'] & .
\end{tikzcd}$$

The precise significance of this property will become clear later on; for the time being, it might be amusing to find out how an isomorphism of categories

$$[\mathbf{A} \times \mathbf{B}, \mathbf{C}] \cong [\mathbf{A}, [\mathbf{B}, \mathbf{C}]]$$

can be deduced from the above theorem. Moreover, the diagonal functor $\Delta : \mathbf{B} \to [\mathbf{A}, \mathbf{B}]$ from Example 3.3 can be obtained by applying the theorem to the projection functor $\mathbf{B} \times \mathbf{A} \to \mathbf{B}$.

4 Adjoint Functors

It is now time to unify the various 'unique factorization properties' that have appeared in the previous sections. In fact, the general concept can be distilled easily from the examples:

Definition 4.1 Let $G : \mathbf{A} \to \mathbf{B}$ be a functor. A *G-structured arrow* consists of an object A in $\mathbf{A}$ and a $\mathbf{B}$-morphism $f : B \to GA$; somewhat imprecisely, we will refer to such a structured morphism just as $f : B \to GA$. Given an object B in $\mathbf{B}$, a *G-universal arrow* for B is a G-structured morphism $\eta_B : B \to GA_B$ such that, for each G-structured morphism $f : B \to GC$, there exists a unique $\mathbf{A}$-morphism $f^{\#} : A_B \to C$ such that $Gf^{\#}\eta_B = f$:

$$\begin{tikzcd} B \arrow[r, "\eta_B"] \arrow[dr, "f"'] & GA_B \arrow[d, dotted, "Gf^{\#}"] & A_B \arrow[d, dotted, "f^{\#}"] \\ & GC & C \end{tikzcd} .$$

G is called *adjoint* if there exists a G-universal arrow for each $B \in \mathrm{Ob}\,\mathbf{B}$.

Example 4.2 A subcategory $\mathbf{B}$ of a category $\mathbf{A}$ is reflective iff the inclusion functor $\mathbf{B} \hookrightarrow \mathbf{A}$ is adjoint. In this case, the universal arrows are just the reflections.

Example 4.3 Equivalences (Section 2) are adjoint.

Example 4.4 The forgetful functor $U : \mathbf{Grp} \to \mathbf{Set}$ is adjoint. For a set X, a U-universal arrow is given by the insertion of generators $X \hookrightarrow UFX$, where FX denotes the free group over X (cf. Example 2.11). In a much more trivial way, the forgetful functor $V : \mathbf{Top} \to \mathbf{Set}$ is also adjoint: a universal arrow for a set X is given by the identity map $X \to VDX$, where DX denotes the discrete space with carrier set X. These examples are typical for categories of algebraic and topological nature, respectively.

Example 4.5 Let $\mathbf{A}$ be a category with coproducts. Then, for each object A in $\mathbf{A}$, the functor $hom(A, \_) : \mathbf{A} \to \mathbf{Set}$ (Example 2.7) is adjoint. A $hom(A, \_)$-universal arrow for a set X is given by $\eta_X : X \to hom(A, X \cdot A)$, where $X \cdot A$ denotes the coproduct of an X-indexed family of copies of A and $\eta_X(x)$ is the coproduct injection $\mu_x : A \hookrightarrow X \cdot A$ associated to x. By Example 3.9, the preceding example contains special cases of this statement.

Example 4.6 Given an abelian group G, one has an *internal hom-functor* $Hom(G, \_) : \mathbf{Ab} \to \mathbf{Ab}$ which maps a further abelian group H to the abelian group of homomorphisms $G \to H$ (with pointwise operations). $Hom(G, \_)$ is adjoint; a $Hom(G, \_)$-universal arrow for an abelian group

H is given by the homomorphism $H \to Hom(G, H \otimes G)$ which maps h to the homomorphism $g \mapsto h \otimes g$.

Example 4.7 A product of a J-indexed family of objects in a category $\mathbf{A}$ is essentially a Δ-co-universal arrow, where $\Delta : \mathbf{A} \to [J, \mathbf{A}]$ denotes the diagonal functor (and J is regarded as a discrete category; cf. Examples 3.3, 3.4). More generally, one might define a *limit* of a *diagram*, i.e., a functor $D : \mathbf{I} \to \mathbf{A}$, as a Δ-co-universal arrow $\Delta A \to D$, where Δ denotes the diagonal functor $\mathbf{A} \to [\mathbf{I}, \mathbf{A}]$. It is instructive to reword this definition in a more explicit way, as done in Section 1 for the case of products; of course, this 'explicit rewording' should really be understood as the primary definition. A category $\mathbf{A}$ is called *complete* if each diagram $\mathbf{I} \to \mathbf{A}$, where $\mathbf{I}$ is a small category, has a limit in $\mathbf{A}$ (such limits are also called *small*).

For example, take $\mathbf{I} = \bullet \rightrightarrows \bullet$. A limit of a diagram $\mathbf{I} \to \mathbf{A}$, i.e., of a pair of parallel morphisms $f, g : A \to B$ in $\mathbf{A}$, is essentially a morphism $e : E \to A$ such that $fe = ge$ and, whenever $fh = gh$ for some morphism $h : C \to A$, there exists a unique morphism $\bar{h} : C \to E$ such that $e\bar{h} = h$.

$$\begin{tikzcd} E \arrow[r, "e"] & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \\ C \arrow[u, dotted, "\bar{h}"] \arrow[ur, "h"'] & & \end{tikzcd} .$$

e is called the *equalizer* of f and g. What are limits of diagrams $\mathbf{I} \to \mathbf{A}$, where $\mathbf{I}$ is $\begin{tikzcd} & \bullet \arrow[d] \\ \bullet \arrow[r] & \bullet \end{tikzcd}$ or $\emptyset$?

Remark 4.8 Equalizers as defined in the above example are easily seen to be extremal monomorphisms. Thus, an equalizer of a pair (f, g) intuitively represents the 'largest subobject where f and g coincide'; e.g., in **Set**, it is just the subset of elements x such that $f(x) = g(x)$. All small limits can be constructed from products and equalizers, provided that these exist [1]; in particular, categories with products and equalizers are complete. This applies to most of the categories presented so far; e.g., **Set**, **Grp** and **Top** are all complete.

Adjoint functors can be defined in several equivalent ways. To begin, note that an adjoint functor $G : \mathbf{A} \to \mathbf{B}$ gives rise to a functor $F : \mathbf{B} \to \mathbf{A}$: One can define F on objects by $FB = A_B$, where $\eta_B : B \to GA_B$ is a G-universal arrow, and on morphisms $f : B \to C$ by commutation of

$$\begin{tikzcd} B \arrow[r, "\eta_B"] \arrow[d, "f"'] & GFB \arrow[d, dotted, "GFf"] \\ C \arrow[r, "\eta_C"] & GFC \end{tikzcd} .$$

By construction, we simultaneously obtain a natural transformation $\eta : id_{\mathbf{B}} \to GF$. Finally, we can define a morphism $\varepsilon_A : FGA \to A$ for each $A \in \mathrm{Ob}\,\mathbf{A}$ by commutation of

$$\begin{tikzcd} GA \arrow[r, "\eta GA"] \arrow[dr, "id_{GA}"'] & GFGA \arrow[d, dotted, "G\varepsilon_A"] \\ & GA \end{tikzcd},$$

thus obtaining a natural transformation $\varepsilon : FG \to id_{\mathbf{A}}$. By construction, $G\varepsilon\eta_G = id_G$. Moreover, it can be checked that $\varepsilon_F F\eta = id_F$.

Now it turns out that one can define adjoint functors G in terms of the existence of a functor F and natural transformations η and ε as above that satisfy these two equations; the components of η are, then, automatically universal arrows. In such a situation, we say that G is *adjoint to* F; η is called the *unit* and ε the *co-unit* of the adjunction. Note the duality: if G is adjoint to F, then F is co-adjoint to G.

Example 4.9 The forgetful functor $U : \mathbf{Grp} \to \mathbf{Set}$ is adjoint to the free group functor F from Example 2.11. The unit is the natural transformation from Example 3.5. The co-unit is given by the maps $\varepsilon_G : FUG \to G$ that evaluate an element of FUG, i.e., a 'term' (more precisely, an equivalence class of terms) that has elements of G as variables, in G.

Example 4.10 The dual space functor $\_^* : k\text{-}\mathbf{Vec} \to k\text{-}\mathbf{Vec}^{op}$ (Example 3.2) is adjoint to 'itself', i.e., to the functor $(\_^*)^{op} : k\text{-}\mathbf{Vec}^{op} \to k\text{-}\mathbf{Vec}$. The associated unit η is the natural transformation $id_{k\text{-}\mathbf{Vec}} \to \_^{**}$ of Example 3.2; the co-unit is 'the same' transformation, i.e., η regarded as a natural transformation $(\_^{**})^{op} \to id_{k\text{-}\mathbf{Vec}^{op}}$ in $[k\text{-}\mathbf{Vec}^{op}, k\text{-}\mathbf{Vec}^{op}]$.

Finally, the fact that G is adjoint to F can also be expressed by requiring that the functors $hom(F\_,\_)$ and $hom(\_,G\_) : \mathbf{B}^{op} \times \mathbf{A} \to \mathbf{Set}$ be naturally isomorphic (in this case, the universal arrows are obtained as the images of the identities id_{FB}). This formal similarity to adjointness of linear operators originally spawned the name 'adjoint functor' (cf. the historical discussion in [31]).

Adjoint functors have many pleasant properties, especially in connection with further axioms. (In particular, the various definitions of 'algebraic functors' either explicitly include or imply adjointness; see e.g., [1].) An important property shared by all adjoint functors is the fact that they preserve limits (Example 4.7). This statement, which itself is easy to show, has nontrivial partial converses, e.g., Freyd's Adjoint Functor Theorem:

Theorem 4.11 *Let $G : \mathbf{A} \to \mathbf{B}$ be a functor, where $\mathbf{A}$ is complete (Example 4.7). Then G is adjoint iff G preserves limits and satisfies the* solution set condition*: for each object B in* $\mathbf{B}$*, there exists a* solution set $\mathcal{S}$ *of G-structured morphisms such that each G-structured morphism $f : B \to GA$ factors as $f = Gag$, where $g : B \to GC$ belongs to $\mathcal{S}$ and $a : C \to A$ is an* $\mathbf{A}$*-morphism.*

(Note that G-universal arrows 'are' singleton solution sets.)

There are several variants of this theorem that replace the somewhat elusive solution set condition by more manageable conditions, at the price of imposing stronger assumptions on the involved categories. E.g., assume that $\mathbf{A}$ as in the above theorem is *well-powered*, i.e., for each object A in $\mathbf{A}$, there exists, up to isomorphism, only a set of monomorphisms ('subobjects') $B \to A$, and has a *cogenerator*, i.e., an object C such that, whenever $f, g : A \to B$ are morphisms such that $hf = hg$ for each morphism $h : B \to C$, then $f = g$. Then G as in the theorem is adjoint iff G preserves limits. This statement can be applied, e.g., to show that **HComp** (Example 1.17) is reflective in **Top** (which is, of course, well known, but nevertheless not entirely trivial): By virtue of Tychonoff's Theorem (and the fact that equalizers in **Top** of maps between Hausdorff spaces are closed subspaces), **HComp** is complete, and the inclusion functor **HComp** $\hookrightarrow$ **Tych** preserves limits (cf. Remark 4.8). Well-poweredness of **HComp** is easily checked, and by Urysohn's Lemma, $[0, 1]$ is a cogenerator in **HComp**.

5 Factorization Systems

A rather different, but equally fundamental type of structure on categories, is represented by the so-called factorization systems. These structures can be thought of as modeling the (related) concepts of subobject and image. The usefulness of an abstract treatment of these concepts lies partly in the fact that there are different notions of subobject or quotient, in general categorical terms as well as, at times, within a fixed category, that will be considered appropriate for various purposes. Recall, for example, the two types of monomorphism from Definition 1.10/Example 1.11 (there are many more; cf. [1]), or compare topological notions of subobject (injective map, subspace embedding, closed embedding). These concepts are unified by the insight that the 'correct' notion of subobject depends on a choice of 'canonical' factorizations of morphisms.

The primordial example is found, as usual, in the category of sets, where any map factors through its image, i.e., as a surjection followed by an injection:

$$\begin{tikzcd} X \arrow[rr, "f"] \arrow[dr] & & Y \\ & f[X] \arrow[ur, hook] & \end{tikzcd} .$$

In the attempt to formally capture this situation, several observations can be made, such as that this factorization is essentially unique in the appropriate sense (which?) or that surjections (or injections) are closed under composition. However, the following definition has turned out to be the most concise and powerful one:

Definition 5.1 A *factorization system* (for morphisms) on a category **A** is a pair $(\mathcal{E}, \mathcal{M})$ of classes of **A**-morphisms such that

- $\mathcal{E}$ and $\mathcal{M}$ are closed under composition with isomorphisms;
- every morphism f in **A** factors as $f = me$, where $m \in \mathcal{M}$ and $e \in \mathcal{E}$;
- **A** has unique $(\mathcal{E}, \mathcal{M})$-*diagonals*, i.e., each commutative square $mf = ge$, where $m \in \mathcal{M}$ and $e \in \mathcal{E}$, admits a unique *diagonal* d such that $de = f$ and $md = g$:

$$\begin{tikzcd} \bullet \arrow[r, "e"] \arrow[d, "f"'] & \bullet \arrow[d, "g"] \arrow[dl, "d"'] \\ \bullet \arrow[r, "m"'] & \bullet \end{tikzcd} .$$

Other standard properties, such as uniqueness of the factorization, closedness of $\mathcal{E}$ and $\mathcal{M}$ under composition, or the fact that the intersection $\mathcal{E} \cap \mathcal{M}$ consists precisely of the isomorphisms in **A**, can be derived from these axioms. In particular, it can be shown that $\mathcal{E}$ and $\mathcal{M}$ determine each other via the diagonalization requirement; i.e., a morphism e belongs to $\mathcal{E}$ iff **A** has $(\{e\}, \mathcal{M})$-diagonals (and dually).

Example 5.2 On every category, there are two trivial factorization systems, where one of the classes contains precisely the isomorphisms and the other all morphisms.

Example 5.3 As indicated above, taking $\mathcal{E}$ as the class of surjections and $\mathcal{M}$ as the class of injections, one defines a factorization system on **Set**. Similarly, surjective and injective homomorphisms form a factorization system on **Grp**. On **Top**, there are many factorization systems that 'lift' the factorization system on **Set**. E.g., quotients and injections form a factorization system on **Top**, and so do surjections and subspace embeddings.

Example 5.4 On **Top**, there is a further natural factorization system $(\mathcal{E}, \mathcal{M})$, where $\mathcal{E}$ does not consist of epimorphisms: take $\mathcal{M}$ as the class of closed subspace embeddings and $\mathcal{E}$ as the class of dense maps.

Example 5.5 On **Cat**, there is a factorization structure $(\mathcal{E}, \mathcal{M})$, where $\mathcal{M}$ consists of the faithful functors and $\mathcal{E}$ consists of the surjective functors that are bijective on objects. In this case, the elements of $\mathcal{M}$ in general fail to be monomorphisms.

Remark 5.6 In 'sufficently complete' categories, certain factorization systems come for free. E.g., in categories with *intersections* and finite limits, (extremal epimorphisms, monomorphisms) is a factorization system, and in categories with intersections and equalizers, epimorphisms and extremal monomorphisms form a factorization system. Here, an *intersection* of a family of monomorphisms ('subobjects') $m_i : B_i \to A$, indexed over a class I, is a limit of the obvious associated diagram.

Remark 5.7 Factorization structures are closely related to full reflective subcategories on the one hand and to *closure operators* (cf. [8, 11, 12]) on the other hand. In fact, under weak additional conditions, these structures are in one-to-one correspondence [3, 7]. One of these additional assumptions is the existence of a *terminal object*, i.e., an object 1 such that there exists a unique morphism $A \to 1$ for each object A. The correspondence between reflective subcategories and factorization systems associates to a factorization system $(\mathcal{E}, \mathcal{M})$ the full subcategory of objects A such that the unique morphism $A \to 1$ belongs to $\mathcal{M}$, and to a full reflective subcategory with *reflector* R (i.e., the embedding is adjoint to R) a factorization structure $(\mathcal{E}, \_)$, where $\mathcal{E}$ is the preimage of the class of isomorphisms under R.

Moreover, factorization structures play an important role in determining whether or not full subcategories are reflective [20, 26], in particular subcategories defined by abstract 'axioms' in a certain sense (*orthogonal* subcategories) [16].

The next step onward leads to factorization systems for *sources*, i.e., (class-indexed) families of morphisms with a common domain. Such systems consist of a class $\mathcal{E}$ of morphisms and a *conglomerate* ('very large collection'; [1]) $\mathbf{M}$ of sources. The factorization property requires that each source $(f_i : A \to B_i)_I$ factors as $f_i = m_i e$ $(i \in I)$, where $e \in \mathcal{E}$ and the source $(m_i)_I$ belongs to $\mathbf{M}$:

$$
\begin{tikzcd}
A \arrow[rr, "f_i"] \arrow[dr, "e"'] & & B_i \\
& \bullet \arrow[ur, "m_i"'] &
\end{tikzcd}
.
$$

The diagonalization condition is analogous to the one in Definition 5.1; of course, the right-hand side of the commutative square now has to be replaced with a source. It is interesting to note that, in the case of sources, the uniqueness requirement in the diagonalization condition can be left out, since $\mathcal{E}$ can be shown to consist of epimorphisms without using uniqueness of diagonals [1].

Example 5.8 The factorization systems for morphisms from Example 5.3 extend to factorization systems for sources: e.g., surjective maps and *point-separating* sources, i.e., sources $(m_i)_I$ such that $m_i(x) = m_i(y)$ for all i implies $x = y$, form a factorization system for sources on **Set**, similarly for **Grp**.

Example 5.9 The factorization system (dense maps, closed embeddings) on **Top** does not extend to sources, since dense maps in general fail to be epimorphisms in **Top**. Its restriction to **Haus**, however, does extend to sources. This is best understood by means of general methods [1], since the relevant class of sources does not really seem to lend itself to a tangible description other than the one that arises from the abstract construction.

One interesting, if simple, application of factorization systems for sources deals with reflective subcategories: Let $(\mathcal{E}, \mathbf{M})$ be a factorization system for sources on a category **A**. Call a full reflective subcategory **B** of **A** *$\mathcal{E}$-reflective* if the associated reflections belong to $\mathcal{E}$. Then a full isomorphism-closed subcategory **B** of **A** is $\mathcal{E}$-reflective iff **B** is *closed under* **M**-*sources* in **A**, i.e., whenever a source $(m_i : A \to B_i)_I$ belongs to **M** and the B_i are **B**-objects, then A is a **B**-object. This condition is often easy to check; e.g., using factorization systems mentioned above, one immediately verifies (again) that **Ab** is reflective in **Grp** and that **Haus** and the full subcategory spanned by the completely regular spaces are reflective in **Top**.

Moreover, by virtue of this characterization, every full subcategory of **A** is contained in a smallest $\mathcal{E}$-reflective subcategory, its *$\mathcal{E}$-reflective hull*. This contrasts sharply with the situation in the case of arbitrary full reflective subcategories; e.g., intersections of two full reflective subcategories need not be reflective even in well-behaved categories like **Top** [37].

Remark 5.10 There are further useful generalizations of factorization systems: factorizations of *structured sources*, i.e., sources that consist of structured morphisms w.r.t. some fixed functor, are important tools in categorical algebra and categorical topology [1], in particular as a means of expressing properties of functors. E.g., a functor is adjoint iff it admits some factorization structure for structured sources. *Essentially algebraic* functors are (roughly) defined as functors that admit suitable factorizations and reflect isomorphisms [1]; this definition does indeed to a certain extent capture the algebraic character of categories.

Factorizations of *flows*, i.e., class-indexed families of set-indexed sources with a common 'family of codomains', have been investigated in [23]. Remarkably, the theorems of the form 'sufficiently complete categories possess factorization systems' that one has in the case of factorization systems for morphisms or sources [1] have actual converses in this extended setting.

6 Cartesian Closedness

The starting point for one of the most fruitful concepts of category theory is the following almost trivial observation: a function f that takes arguments in two sets X and Y is essentially the same as a function that assigns to each $x \in X$ a function that takes one argument in Y; specifically, the latter function maps $y \in Y$ to $f(x, y)$. Logicians call this process 'currying' (in honor of *Haskell B. Curry*). In the language of category theory, this fact can be expressed by saying that

$$hom(X \times Y, Z) \cong hom(X, hom(Y, Z))$$

or, noticing furthermore that this isomorphism is 'natural in the variables X and Z', that one has a natural isomorphism $hom(\_ \times Y, \_) \cong hom(\_, hom(Y, \_))$. In other words: for each set Y, the functor $\_ \times Y : \mathbf{Set} \to \mathbf{Set}$ is co-adjoint to $hom(Y, \_)$ (cf. Section 4).

The situation becomes less trivial when considered in other categories. E.g., the isomorphism as above does not work for vector spaces, since the elements of $L(V, L(W, U))$ (where $L(V, W)$ denotes the space of linear maps from V to W) correspond to *bilinear* rather than linear maps $V \times W \to U$ (see also Example 4.6). In **Top**, the problem is even more intriguing, since it is a priori unclear which topology the set of continuous maps between two spaces should carry, respectively whether there exists a suitable topology at all.

It is now time to explicitly state

Definition 6.1 A category $\mathbf{A}$ is called *cartesian closed* if it has finite products, and the functors $\_ \times A : \mathbf{A} \to \mathbf{A}$ ($A \in \operatorname{Ob} \mathbf{A}$) are coadjoint. In this case, the corresponding co-universal arrows are written in the form

$$ev : B^A \times A \to B;$$

B^A is called a *power object*, and ev is called the *evaluation morphism.*

An important consequence of cartesian closedness is that the functors $\_ \times A$ preserve colimits (cf. Section 4), in particular coequalizers. This shows, e.g., that **Top** indeed fails to be cartesian closed: given a (Hausdorff) space X, $\_ \times X$ preserves quotient maps iff X is core (locally) compact [10].

Example 6.2 **Set** is cartesian closed. The power objects Y^X are the sets of maps $X \to Y$, and the evaluation map sends (f, x) to $f(x)$.

Example 6.3 Let **Rel** denote the category of binary relations, i.e., of objects (X, ρ), where X is a set and $\rho \subset X \times X$; a map $f : (X, \rho) \to (Y, \sigma)$ is a morphism if $f^2[\rho] \subset \sigma$. **Rel** is cartesian closed; power objects $(Y, \sigma)^{(X, \rho)}$ consist of *maps* $X \to Y$, where two maps $f, g : X \to Y$ are in relation iff $(x_1, x_2) \in \rho$ implies $(f(x_1), g(x_2)) \in \sigma$.

An interesting comparison is provided by the full subcategory **ReRe** of **Rel** spanned by the reflexive relations. **ReRe** is also cartesian closed; however, the power objects $(Y, \sigma)^{(X,\rho)}$ now consist of *relation morphisms* $(X, \rho) \to (Y, \sigma)$, with the relation on the power object defined as above.

Example 6.4 The fact that **Top** fails to be cartesian closed is often considered a defect in need of repair. This can be achieved by moving on to larger categories, such as the category of pseudotopological spaces [9]; there are, in fact, general methods for extending ('topological') categories in order to make them cartesian closed [22].

Alternatively, one may restrict the class of spaces. E.g., the category **kTop** of k-spaces, i.e., Hausdorff quotients of locally compact spaces (equivalently: compactly generated Hausdorff spaces), is cartesian closed (and coreflective in **Haus**). Note, however, that finite products of k-spaces (even of two identical ones) in **Haus** and, hence, function spaces when endowed with the compact-open topology may fail to be k-spaces [15]; products and power objects in **kTop** are formed by coreflection.

Example 6.5 The category **Cat** of small categories is cartesian closed; the power objects are the functor categories (this finally explains the significance of Theorem 3.10).

Example 6.6 It is easy to see that **Grp** is not cartesian closed (the functors $\_ \times G$ in general fail to preserve coproducts, i.e., free products). However, the full subcategory **Grpd** of **Cat** spanned by the small groupoids is easily seen to be closed under products and functor categories, and hence is cartesian closed. This is one of the main reasons why groupoids, which at first sight hardly seem to differ from groups at all (e.g., every connected groupoid, i.e., every groupoid in which all hom-sets are nonempty, is equivalent to a one-object groupoid, i.e., a group), are a useful generalization.

It is interesting to observe why the usual way to form 'function groups' (namely, by equipping the set of homomorphisms between two groups with the pointwise operations) fails to produce power objects: here, the homomorphisms are taken as elements of a group, i.e., as *morphisms* of a one-object groupoid, while they should more 'correctly' be viewed as *objects* of a groupoid.

Example 6.7 Heyting algebras, when regarded as categories in the same way as ordered sets (Section 1), are cartesian closed; the 'power objects' are just the relative pseudocomplements.

Having been recognized as a fundamental property of the category of sets, cartesian closedness plays a central role in the categorical abstraction of set theory, the theory of *topoi*; see Section 7. One of the intriguing aspects of that theory is its ability to capture and describe features of a category within the category itself. Simple examples work already for cartesian closed categories; e.g., in a cartesian closed category **A**, we have not only a representation of morphisms $A \to B$ as elements $1 \to B^A$ (where 1 denotes the terminal object; cf. Remark 5.7), but also a representation of the composition operator as a family of morphisms $comp : C^B \times B^A \to C^A$, defined by commutation of the diagram

$$\begin{tikzcd} C^A \times A \arrow[rr, "ev"] & & C \\ C^B \times B^A \times A \arrow[u, "comp \times id_A"] \arrow[r, "id_{C^B} \times ev"'] & C^B \times B . \arrow[ur, "ev"'] & \end{tikzcd}$$

Remark 6.8 Another area of application that adds importance to the theory of cartesian closed categories is the semantics of (typed) λ-calculi: given a typed λ-calculus, one can interpret its types as objects of a given cartesian closed category, with product types as finite products and function types as power objects; terms are interpreted as morphisms. λ-abstraction then just corresponds to applying the co-universal property of the power objects.

In fact, λ-calculi are even in a certain sense equivalent to cartesian closed categories [28]. The correspondence works roughly as follows: one can construct a cartesian closed category from a typed λ-calculus by taking types as objects and terms as morphisms; conversely, a cartesian closed category gives rise to a λ-calculus which has the objects as types and the morphisms as function symbols.

7 Outlook

There are, of course, numerous topics in category theory, amusing and insightful at the same time, that we have not even touched upon. We shall spend the final part of the tour on just throwing a few key words:

Monads

An adjoint pair of functors $G : \mathbf{A} \rightleftarrows \mathbf{B} : F$ with unit η and co-unit ε gives rise to certain operations on, say, the 'base' category $\mathbf{B}$, in particular a functor $T := GF : \mathbf{B} \to \mathbf{B}$ and natural transformations $\eta : id_{\mathbf{B}} \to GF = T$ and $\mu := G\varepsilon_F : T^2 \to T$ that fulfill certain equations. Abstracting from the adjoint pair, such a set of data (T, η, μ) is called a *monad* (also *triple* or *standard construction*).

It has turned out that each such monad is induced by an adjoint pair in the way just described. Indeed, there are several 'canonical constructions' for such adjoint pairs, in particular a 'smallest' one, called the *Kleisli category* [27] (which more recently has assumed importance in the design of programming languages [32]), and a 'largest' one, the *Eilenberg–Moore category* [14]. The latter consists of objects called algebras, and indeed typical algebraic categories (together with their forgetful functors) such as those of groups, rings, or monoids are isomorphic to the Eilenberg–Moore category of the associated monad on **Set**. The study of the relationship of adjoint functors to the associated categories of algebras has led to many interesting results.

Algebraic Theories

An alternative way to capture 'general algebra' in a categorical setting is provided by the observation, due to Lawvere [29], that one can replace the rather cumbersome signatures of operation symbols used in classical universal algebra by certain small categories (with additional properties such as, in particular, existence of finite products) called *algebraic theories* that represent operations as morphisms and equations via equality of composites. A model of such a theory $\mathbf{C}$ (i.e., in the classical setting, an implementation of the symbols as actual operations that respect the given equations) is then just a product preserving functor $\mathbf{C} \to \mathbf{Set}$, and homomorphisms between the models are indeed natural transformations. Having thus abstracted away the concrete syntax, one can perform arguments in a rather simpler and more elegant fashion than in the classical setting; for a recent example see [33].

Locally Presentable Categories

A third theory of 'algebraic' categories differs from the preceding ones both in covering a broader range of (finitary) algebraic structures and in considering categories independently of a choice of forgetful functor. In this case, the fundamental observation states that in many 'algebraic' categories, all objects can be 'incrementally constructed' via so-called *directed colimits* from a representative set of objects with a finite presentation (by generators and relations), and that these finitely presented objects can up to a certain degree be characterized by the fact that the associated hom-functors preserve the said directed colimits. Categories in which the abstract requirements distilled from these heuristics are satisfied (this includes, e.g., **Cat**) are called *locally (finitely) presentable* [17]. The theory that has unfolded from these comparatively simple axioms is astonishingly extensive; see e.g., [2].

Topological Functors

The development of an abstract theory of topological categories is more recent. As the crucial concept, a generalization of the notion of initial topology (w.r.t. a source) has emerged; essentially, an initial source (w.r.t.

a fixed forgetful functor) in this general sense is a source that can be cancelled from statements of the form ' ... is a morphism' (to be used in the same way as one would say that a *map* between topological spaces is or is not continuous). A functor F is called *topological* if each F-structured source admits a (unique) initial lift; e.g., the usual product topology is just the initial lift of the structured source consisting of the projections. This is the starting point of categorical topology, which has provided a unifying background for general topology. Topological functors have numerous pleasant properties; moreover, many categories (more precisely: functors) admit elegantly constructed topological completions; see e.g., [21, 24].

Topoi

Surprisingly, one of the most successful branches of category theory is concerned with a generalized treatment not of complicated algebraic or topological structures, but of set theory. Here, set theory is, of course, not regarded as dealing with a forever fixed universe of sets, but rather as depending on the particularities of the employed logic, axiomatization and other factors. Under the basic assumption that the usefulness of a collection of 'sets' hinges primarily on the constructions that can be performed with these 'sets', a set of axioms for 'sufficiently **Set**-like' categories has evolved which defines the notion of *topos.*

To wit, a topos is a cartesian closed category with certain additional properties which are inspired by typical properties of **Set**; most notable among these is the existence of *representations* in a certain sense for partial morphisms, i.e., morphisms defined only on a subobject of the domain [1, 18, 25]. Typical topoi, besides **Set**, are categories of sheaves; in fact, the early development of topos theory was largely motivated by the needs of algebraic geometry [19]. Topos theory has acquired particular importance due to its close relation to constructive logic.

In topology, the concept of topos has been shown to be too strong [38]. However, weakening the representation requirement to cover only partial morphisms defined on *extremal* subobjects, one arrives at the notion of *quasitopos*, which does cover interesting topological categories. In particular, the category of pseudotopological spaces is a quasitopos. Moreover, there are quasitopoi in computer science, such as the category of projection spaces [13].

Acknowledgement

The diagrams in this paper were typeset using Paul Taylor's diagram package.

References

[1] J. Adámek, H. Herrlich and G.E. Strecker. *Abstract and Concrete Categories*, Wiley Interscience, New York, 1990.

[2] J. Adámek and J. Rosický. *Locally Presentable and Accessible Categories*, Cambridge University Press, 1994.

[3] F. Borceux. *Handbook of Categorical Algebra 1*, Cambridge University Press, 1994.

[4] N. Bourbaki. Éléments de mathématique, Vol. VII, Algèbre, Livre II, Algèbre multilinéaire, ch. 3, *Actualités Sci. Indust.*, **1044** Hermann, Paris, 1948.

[5] H. Brandt. Über eine Verallgemeinerung des Gruppenbegriffs, *Math. Ann.*, **96** (1926), 360–366.

[6] R. Brown. From groups to groupoids: a brief survey, *Bull. London Math. Soc.*, **19** (1987), 113–134.

[7] C. Cassidy, M. Hébert and G. M. Kelly: Reflective subcategories, localizations and factorization systems, *J. Austral. Math. Soc.*, **38** (1985), 287–329, Corrigenda **41** (1986), 286.

[8] G. Castellini. Categorical closure operators, this volume.

[9] G. Choquet. Convergences, *Ann. Univ. Grenoble*, Sect. Sci. Math. Phys., II, Ser. **23** (1948), 57–112.

[10] B. J. Day and G. M. Kelly. On topological quotient maps preserved by pullbacks or products. *Proc. Cambridge Phil. Soc.*, **67** (1970), 553–558.

[11] D. Dikranjan and E. Giuli. Closure operators I, *Topology Appl.*, **27** (1987), 129–143.

[12] D. Dikranjan and W. Tholen. *Categorical Structure of Closure Operators*, Kluwer Academic Publishers, Dordrecht, 1995.

[13] H. Ehrig and H. Herrlich. The construct PRO of projection spaces: its internal structure, *Lect. Notes Comp. Sci.*, **393**, Springer, Berlin, 1989.

[14] S. Eilenberg and S. Mac Lane. General theory of natural equivalences, *Trans. Amer. Math. Soc.*, **58** (1945), 231–294.

[15] R. Engelking. *General Topology*, Heldermann, Berlin, 1989.

[16] P. J. Freyd and G. M. Kelly. Categories of continuous functors, I, *J. Pure Appl. Algebra*, **2** (1972), 169–191.

[17] P. Gabriel, F. Ulmer. *Lokal präsentierbare Kategorien*, *Lect. Notes Math.*, **221**, Springer, Berlin, 1971.

[18] R. Goldblatt. *Topoi. The Categorial Analysis of Logic*, rev. ed., North-Holland, Amsterdam, 1984.

[19] M. Artin, A. Grothendieck and J. L. Verdier. *Théorie des topos et cohomologie étale des schémas*, *Lect. Notes Math.* **269**, **270**, Springer, Berlin, 1972.

[20] H. Herrlich. *Topologische Reflexionen und Coreflexionen*, *Lect. Notes Math.*, **78**, Springer, Berlin, 1968.

[21] H. Herrlich. Initial completions, *Math. Z.*, **150** (1976), 101–110.

[22] H. Herrlich. Topological improvements of categories of structured sets, *Topology Appl.*, **27** (1987), 145–155.

[23] H. Herrlich and W. Meyer. Factorization of flows and completeness of categories, *Quaest. Math.*, **17** (1994), 1–11.

[24] H. Herrlich and G. E. Strecker. Semi-universal maps and universal initial completions, *Pacific J. Math.*, **82** (1979), 407–428.

[25] P. T. Johnstone. *Topos Theory*, Academic Press, London, 1977.

[26] J. F. Kennison. Full reflective subcategories and generalized covering spaces, *Illinois J. Math.*, **12** (1968), 353–365.

[27] H. Kleisli. Every standard construction is induced by a pair of adjoint functors, *Proc. Amer. Math. Soc.*, **16** (1065), 544–546.

[28] J. Lambek and P. J. Scott. *Introduction to Higher Order Categorical Logic*, Cambridge University Press, 1986.

[29] F. W. Lawvere. *Functorial Semantics of Algebraic Theories*, PhD thesis, Columbia University, 1963. Summary in *Proc. Natl. Acad. Sci. USA*, **50** (1963), 869–872.

[30] F. Linton. Some aspects of equational categories, in *Proceedings of the Conference on Categorical Algebra at La Jolla*, Springer, Berlin, 1966.

[31] S. Mac Lane. *Categories for the Working Mathematician*, Springer, Berlin, 1972.

[32] E. Moggi. Notions of Computation and Monads, *Inform. and Comput.*, **93** (1991), 55–92.

[33] H. E. Porst. Equivalence of varieties in general and of Bool in particular, to appear in *Algebra Universalis.*

[34] L. Schröder. *Composition Graphs and Free Extensions of Categories* (in German), PhD thesis, University of Bremen, 1999; also: Logos, Berlin, 1999.

[35] L. Schröder. Monads on composition graphs. Submitted to *Appl. Cat. Struct.*

[36] L. Schröder and H. Herrlich. Free adjunction of morphisms, *Appl. Cat. Struct.*, **8** (2000), 595–606.

[37] V. Trnkova, J. Adámek and J. Rosický. Topological reflections revisited, *Proc. Amer. Math. Soc.*, **108** (1990), 605–612.

[38] O. Wyler. Are there topoi in topology? *Lect. Notes Math.*, **540**, Springer, Berlin, 1976, 699–719.

Lutz Schröder
Department of Mathematics and Computer Science
University of Bremen, P.O. Box 330440, 28334 Bremen, Germany
E-mail: `lschrode@informatik.uni-bremen.de`

The Functor that Wouldn't be

— A Contribution to the Theory of Things that Fail to Exist —

Y.T. Rhineghost[1]

Key words: Injective hull.

AMS Subject Classifications (2000): 18A22, 18B05.

S. Masta, do you have a problem for me — please?

M. What kind do you want?

S. Oh, a mathematical one — please!

M. Well, you know that in the category **Set** of sets the injectives are well behaved, i.e., every set X has an injective hull $j_X \colon X \longrightarrow X^*$. Are these arrows *reflection arrows*? Or, more generally:

B.T.B.[2]

Problem 1:	Does the inclusion functor $E\colon \mathbf{Inj} \longrightarrow \mathbf{Set}$ from the full category **Inj** of **Set**, consisting of injective (= nonempty) objects, have a coadjoint?

S. Oh! That's easy. The j_X's are reflection arrows. So the answer is Yes.

M. Boy! Are you sure you didn't overlook something?

> "Right now the Kane brain
> was full of a disturbing intuition
> that something was very wrong."
>
> From *Brand of Kane* by Hugh B. Cave

---

[1]The author acknowledges virtual support by the Society for the Preservation of Nonexistent Things and moral support by the Friends of Unnatural Problems.

[2]From H. Belloc: *Cautionary Tales* (1907).

After 5 minutes:

S. Well — maybe the empty set. Oh, yes, $j_\emptyset\colon \emptyset \longrightarrow \emptyset^*$ fails to be a reflection arrow; though all the other j_X's are, and $j_\emptyset$ is a weak reflection arrow[3].

M. So, what is the answer to my question?

S. It boils down to finding out whether there is a reflection arrow for the empty set. — Well, there is none: no map $\emptyset \longrightarrow Y$ from $\emptyset$ into a non–empty set Y is a reflection arrow — though all of them happen to be weak reflection arrows[3]. So the answer to your question is No. — Puh! That was more tricky than I thought!

— Masta, do you, perhaps, have a simpler problem for me — please?

M. Well, since there is no coadjoint $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ for the inclusion functor $E\colon \mathbf{Inj} \longrightarrow \mathbf{Set}$, you may ask for less: Are injective hulls in **Set** at least *natural*? I.e.,

Problem 2: Does there exist a functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ such that

(a) $FX = X^*$ for each set X,

(b)
$$\begin{tikzcd} X \arrow[r, "j_X"] \arrow[d, "f"'] & FX \arrow[d, "Ff"] \\ Y \arrow[r, "j_Y"'] & FY \end{tikzcd}$$

commutes for each map $f\colon X \longrightarrow Y$, where the $j_X\colon X \longrightarrow X^*$ form a collection of injective hulls?

An hour later:

S. Masta! Masta! I got it! This time I didn't forget the empty set. The answer to your problem is No, and the following proves this. Assume for simplicity that

$$j_X\colon X \longrightarrow X^* = \begin{cases} j_\emptyset\colon & \emptyset \longrightarrow \{0\}, \quad \text{if } X = \emptyset, \\ \mathrm{id}_X\colon & X \longrightarrow X, \quad \text{otherwise.} \end{cases}$$

---

[3]i.e., for each map $\emptyset \xrightarrow{f} X$ with $X \in \mathbf{Inj}$ there exists a (not necessarily unique) map $\emptyset^* \xrightarrow{f^*} X$ with $f = f^* \circ j_\emptyset$.

Assume further that there would be a functor F with properties (a) and (b) above. Let f and g be the two different maps from $\{0\}$ to $\{1,2\}$, and define $h = f \circ j_\emptyset = g \circ j_\emptyset$. Then (b) would imply that:

$$Fj_\emptyset = \mathsf{id}_{\{0\}}, \ Ff = f, \text{ and } Fg = g.$$

Thus $f = f \circ \mathsf{id}_{\{0\}} = Ff \circ Fj_\emptyset = F(f \circ j_\emptyset) = F(h) = F(g \circ j_\emptyset) = Fg \circ Fj_\emptyset = g \circ \mathsf{id}_{\{0\}} = g$.

A contradiction. Hurray!

Masta, this was quite tough. Would you be so nice to give me another problem; a simpler one — please?

M. This was a cute proof. Moreover, you have shown that the inclusion functor $E\colon \mathbf{Inj} \longrightarrow \mathbf{Set}$ has no left inverse F with the additional property $F\emptyset = \{0\}$. So it seems natural to ask the following:

Problem 3:	Does there exist a left inverse $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ for the inclusion functor $E\colon \mathbf{Inj} \longrightarrow \mathbf{Set}$?

Ten minutes later:

S. Masta! Masta! This was easy! The answer, again, is No. Assume that F would be a left inverse for E. Define the maps $j_\emptyset, f, g$, and h as before. Then $Ff = f$, $Fg = g =$ and

$$Ff \circ Fj_\emptyset = F(f \circ j_\emptyset) = Fh = F(g \circ j_\emptyset) = Fg \circ Fj_\emptyset$$

as before. Moreover $F\emptyset \neq \emptyset$ and $F\{0\} = \{0\}$ would imply, as my friend Lutz observed, that the map $Fj_\emptyset\colon F\emptyset \longrightarrow F\{0\}$ would be surjective, and thus the above equations would imply

$$f = Ff = Fg = g,$$

a contradiction again.

Masta! Do you have another problem for me?

M. Boy, since the above natural questions all have negative answers, you may want to investigate the following — perhaps somewhat *unnatural* problem — requiring much less from the desired functor:

Problem 4:	Does there exist a functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ with $FX = X$ for each non–empty set X?

Two days later:

S. Masta! I have it! This was tough! The answer surprised me a = lot. It is No again. The proof required more time. Assume that there would be a functor $G\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ with $GX = X$ for each non–empty set X. Since for each non–empty set X, the map $\emptyset_X\colon \emptyset \longrightarrow X$ factors through $\{0\}$, the map $G\emptyset_X\colon G\emptyset \longrightarrow GX$ would factor through $G\{0\} = \{0\}$:

$$\begin{tikzcd}
\emptyset \arrow[rr, "\emptyset_{\{0\}}"] \arrow[dr, "\emptyset_X"'] & & \{0\} \arrow[dl, "k"] \\
& X &
\end{tikzcd}
\qquad
\begin{tikzcd}
G\emptyset \arrow[rr, "G\emptyset_{\{0\}}"] \arrow[dr, "G\emptyset_X"'] & & \{0\} = G\{0\} \arrow[dl, "Gk"] \\
& GX &
\end{tikzcd}$$

Thus $G\emptyset_X$ would be constant and, consequently for any function $k\colon \{0\} \longrightarrow X$, $Gk\colon \{0\} \longrightarrow X$ would not depend on the choice of k. Let us call it k_X. Therefore we could define a functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ by

$$FX \quad = \quad \left\{ \begin{array}{ll} \{0\}, & \text{if } X = \emptyset \\ X, & \text{if } X \neq \emptyset \end{array} \right\} \qquad \text{for each set } X, \text{ and}$$

$$F(X \xrightarrow{f} Y) \quad = \quad \left\{ \begin{array}{ll} \mathrm{id}_{\{0\}}, & \text{if } X = Y = \emptyset \\ k_Y, & \text{if } X = \emptyset \neq Y \\ f, & \text{if } X \neq \emptyset \end{array} \right\} \qquad \text{for each map } f.$$

Then each non–empty set X would have a distinguished element, = namely $F\emptyset_X(0)$, and each Ff would preserve this element (so that F could be considered as a functor from **Set** into the category of pointed sets). For each $n \in \mathbb{N}$ consider the set $X_n = \{0, 1, \ldots, n\}$ and assume, without loss of generality, that 0 is the distinguished element of each X_n.

(1) Let f be a constant map. Then f factors through X_0. Thus Ff factors through X_0. Therefore Ff is a constant map with value 0.

(2) Consider the isomorphism $s\colon X_1 \longrightarrow X_1$, defined by $s(0) = 1$ and $s(1) = 0$. Then Fs is an isomorphism with $Fs(0) = 0$. Thus $Fs = \mathrm{id}_{X_1}$.

(3) Consider the inclusion map $f_0\colon X_1 \longrightarrow X_2$. Then $Ff_0(0) = 0$. Moreover, f_0 is a section, hence Ff_0 is a section, thus injective. Assume, without loss of generality, that $Ff_0(1) = 1$. Hence $Ff_0 = f_0$.

(4) Consider the isomorphism $t\colon X_2 \longrightarrow X_2$, defined by $t(0) = 1$, $t(1) = 2$, and $t(2) = 0$. Then Ft is = an isomorphism with $Ft(0) = 0$. Thus Ft is either id_{X_2} or the map g that switches 1 and 2. However, the latter cannot be since $t^3 = \mathsf{id}_{X_2}$, but $g^3 \neq \mathsf{id}_{X_2}$. Thus $Ft == \mathsf{id}_{X_2}$.

(5) Let $f\colon X_1 \longrightarrow X_2$ be an arbitrary injection. Then there exist $g \in \{s, \mathsf{id}_{X_1}\}$ and $h \in \{t, t^2, \mathsf{id}_{X_2}\}$ with $f = h \circ f_0 \circ g$. Thus

$$Ff = Fh \circ Ff_0 \circ Fg = \mathsf{id}_{X_2} \circ f_0 \circ \mathsf{id}_{X_1} = f_0.$$

(6) Consider the isomorphism $f_1\colon X_2 \longrightarrow X_2$, defined by $f_1(0) = 0$, $f_1(1) = 2$, and $f_1(2) = 1$. Then Ff_1 is a bijection with $Ff_1(0) = 0$. The map $f = f_1 \circ f_0\colon X_1 \longrightarrow X_2$ an injection. Thus by (5), we get

$$f_0 = Ff = Ff_1 \circ Ff_0 = Ff_1 \circ f_0.$$

In particular: $1 = f_0(1) = Ff_1(f_0(1)) = Ff_1(1)$. Consequently $Ff_1 = \mathsf{id}_{X_2}$.

(7) Let $f\colon X_2 \longrightarrow X_2$ be a bijection with exactly one fixed point. Then $f \in \{f_1 \circ \mathsf{id}_{X_2},\ f_1 \circ t,\ f_1 \circ t^2\}$. Thus $Ff \in \{Ff_1,\ Ft,\ Ft^2\} = \{\mathsf{id}_{X_2}\}$. So $Ff = \mathsf{id}_{X_2}$.

(8) Consider the map $f_2\colon X_2 \longrightarrow X_1$, defined by $f_2(0) = 0$ and $f_2(1) = f_2(2) = 1$. Then $f_2 \circ f_0 = \mathsf{id}_{X_1}$. Thus $\mathsf{id}_{X_1} = Ff_2 \circ Ff_0 = Ff_2 \circ f_0$, which implies $Ff_2(1) = 1$.

(9) Consider the map $f_3 = f_0 \circ f_2\colon X_2 \longrightarrow X_2$. Then $Ff_3 = Ff_0 \circ Ff_2 = f_0 \circ Ff_2$. Thus $Ff_3(1) = 1$.

(10) Consider the map $f_4\colon X_2 \longrightarrow X_2$, defined by $f_4(0) = f_4(1) = 0$, and $f_4(2) = 1$. Then $f_4 = g \circ f_3 \circ h$, where $h\colon X_3 \longrightarrow X_3$ is the bijection, defined by $h(0) = 2$, $h(1) = 1$, and $h(2) = 0 =$ and $g\colon X_3 \longrightarrow X_3$ is the bijection, defined by $g(0) = 1$, $g(1) = 0$, and $g(2) = 2$. Thus, by (7), $Ff_4 = Fg \circ Ff_3 \circ Fh = Ff_3$. Consequently $F(f_4 \circ f_4)(1) = F(f_3 \circ f_3)(1) = Ff_3(Ff_3(1)) = Ff_3(1) = 1$. On the other hand $f_4 \circ f_4$ is a constant map. Thus, by (1), $F(f_4 \circ f_4)(1) = 0$. Contradiction!

M. This is amazing![4] You have in fact shown that there is no functor F from the full subcategory $\mathbf{A}$ of $\mathbf{Set}$, consisting of the four objects $X_0 = \{0\}$, $X_1 = \{0, 1\}$, $X_2 = \{0, 1, 2\}$, and $\emptyset$, to the full subcategory $\mathbf{B}$, consisting of X_0, X_1, and X_2, that keeps the objects X_0, X_1 and X_2 fixed. This is quite surprising, in view of the fact that

[4]Perhaps not so to those who (like Jirka) are familiar with V. Koubek's paper *Set Functors* in Comment. Math. Univ. Carolinae **12** (1971) 175–195. The results in there provide another (completely different) proof of the fact that Problem 4 has a negative answer.

(a) there does exist a functor from **A** to **B** that keeps X_0 and X_2 fixed, and

(b) there also does exist a functor from the full subcategory of **Set**, consisting of the three objects X_0, X_1, and $\emptyset$, into the one consisting of X_0 and X_1, that keeps X_0 and X_1 fixed.

Nice work, boy!

S. Masta, are there more problems of this kind?

M. There sure are.[5]

Since you have shown that there is no functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ that keeps the objects X_0, X_1 and X_2 fixed, the following unnatural problems arise naturally:

Problem 5: Consider $X_0 = \{0\},\ X_1 = \{0,1\}$ and $X_2 = \{0,1,2\}$. Is there a functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ with any of the following properties:

(a) F keeps X_0 and X_1 fixed?

(b) F keeps X_0 and X_2 fixed?

(c) F has at least two fixed objects?

S. Oh boy! — How strange! Masta, the problems you posed got simpler and simpler in so far as you require less and less from the desired functors; and yet it gets harder and harder to solve them. How can that be?

M. Oh well! Think about it, boy! You will find the answer.

[5] Related problems, — some natural, some unnatural, — some solved, some open —, are contained in the inspiring manuscripts

- B. Banaschewski and J.J.C. Vermeulen: *On the non–functoriality of the maximal ideal space of a commutative ring.* Preprint. Dec. 1998.
- J. Adámek, H. Herrlich, J. Rosický, and W. Tholen: *Injective Hulls are not Natural.* Preprint, Dec. 1999.
- J. Adámek, H. Herrlich, J. Rosický, and W. Tholen: *On a Generalized Small–Object Argument for the Injective Subcategory Problem.* Preprint, May 2000.

The Moral of this priceless work
(If rightly understood)
Will make you – from a little Turk –
Unnaturally good.

From H. Belloc: *The Bad Child's Book of Beasts* (1896)

Y.T. Rhineghost
c/o H. Herrlich
Department of Mathematics
University of Bremen
P.O. Box 33 o4 4o
28334 Bremen, Germany
herrlich@math.Uni-Bremen.de

The Emergence of Functors

— A Continuation of "The Functor that Wouldn't be" —

To be studied in tune with a lazy lackadaisical melody.

Y.T. Rhineghost

Key words: Endofunctor of SET, category of sets, fixpoint, injective hull.

AMS Subject Classifications (2000): 18A22, 18B05.

After Christmas

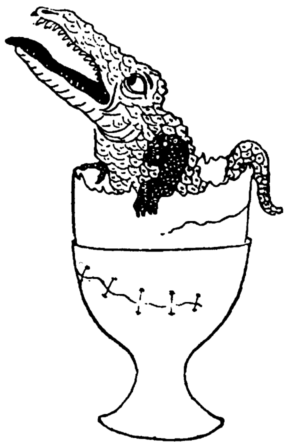
S. Masta — I had a wonderful Christmas!

M. What made it so wonderful?

S. I discovered answers.

M. That sounds great. Let's hear what you got.

S. First, I realized why your problems got harder and harder to solve even though they seemed to get easier and easier.



B.T.B.[1]

M. Why is that so?

S. Masta, simple enough: If the problems get easier, provided their answers are positive, they get harder if their answers are negative, — and the latter is the case with your Problems 1 to 4. There are no functors with the prescribed properties.

M. Fine! But what about Problem 5? Did you solve that?

S. Yes Masta! — I told myself that I am going to crack it even if it harelips

[1]From H. Belloc: *More Beasts for Worse Children (1987).*

all the hogs in Texas. And, by golly! I did: The answer is yes! — Three times yes!

M. That is interesting. Let's see your solution to Problem 5a.

S. Masta — I can construct an endofunctor F of **Set** with the following properties:

(A1) $\lvert F0\rvert = \lvert F1\rvert = 1$,
(A2) $\lvert F2\rvert = 2$,
(A3) $\lvert Fn\rvert = \binom{n}{2} + 1$ for each n with $2 < n <= \infty$,
(A4) $\lvert FX\rvert = \lvert X\rvert$ for each infinite set X,
where $\lvert X\rvert$ is the cardinal number of X, and $n = \{m \in \mathbb{N} \mid m < n\}$ for each natural number n.

M. Go ahead!

S. Masta — the construction will proceed in two steps:

Consider first the functor $G\colon \mathbf{Set} \longrightarrow \mathbf{Set}$ defined by:

(a) $GX = 2^{(2^X)}$,
(b) $G(X \xrightarrow{f} Y) = GX \xrightarrow{Gf} GY$, where $\big(Gf(\alpha)\big)(\beta) = \alpha(\beta \circ f)$ for $\alpha \in GX$ and $\beta \in 2^Y$.

M. Have you checked whether G is a functor?

S. Oh yes — Masta! This is easy, since:

1. $\big(G\,\mathsf{id}_X(\alpha)\big)(\beta) = \alpha(\beta \circ \mathsf{id}_X) = \alpha(\beta)$ for each $\beta \in 2^X$.
   Thus $G\,\mathsf{id}_X(\alpha) = \alpha = \mathsf{id}_{GX}(\alpha)$ for each $= \alpha \in GX$.
   Thus $G\,\mathsf{id}_X = \mathsf{id}_{GX}$.
2. Consider maps $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow Z$. Then $\big(G(g \circ f)(\alpha)\big)(\gamma) = \alpha\big(\gamma \circ (g \circ f)\big) = \alpha\big((\gamma \circ g) \circ f\big) = \big(Gf(\alpha)\big)(\gamma \circ g) = \big(Gg(Gf(\alpha))\big)(\gamma) = \big((Gg \circ Gf)(\alpha)\big)(\gamma)$ for each $\gamma \in 2^Z$.
   Thus $G(g \circ f)(\alpha) = (Gg \circ Gf)(\alpha)$ for each $\alpha \in GX$.
   Thus $G(g \circ f) = Gg \circ Gf$.

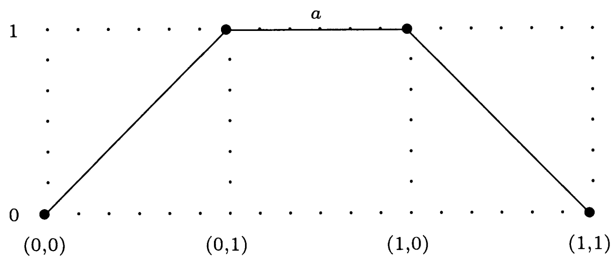
M. Straightforward, indeed. Proceed with step two!

S. Masta — here is the crucial idea: Consider the *subfunctor* F of G determined by

(a) $F\emptyset = \{0_{2^\emptyset}\}$,

(b) $FX = \cup\{Gf[A] \mid f \in X^2\}$ for $X \neq \emptyset$, where $A = \{0_{2^2}, a\}$

with $a\colon 2^2 \longrightarrow 2$ defined by $a(n,m) = n + m(\text{mod } 2)$[2],
while $0_X\colon X \longrightarrow 2$ is the constant map with value 0.
Graph of a:



Then F has the desired properties.

M. Have you checked whether F is a functor?

S. Oh yes — Masta! It suffices to show that

(∗) $\quad Gf[FX] \subset FY \qquad$ for each map $f\colon X \longrightarrow Y$.

But this is obvious unless $X = \emptyset$ and $Y \neq \emptyset$. So it suffices to check (∗) for the empty = map $f\colon \emptyset \longrightarrow Y$. Now for any $\beta \in 2^Y$:

$$\big(Gf(0_{2^\emptyset})\big)(\beta) = 0_{2^\emptyset}(\beta \circ f) = 0 = 0_Y(\beta).$$

Thus $Gf(0_{2^\emptyset}) = 0_Y$. So it remains to be shown that $0_Y \in FY$ for each $Y \neq \emptyset$. For this purpose let $g\colon 2 \longrightarrow Y$ be an arbitrary map. Then:

$$\big(Gg(0_{2^2})\big)(\beta) = 0_{2^2}(\beta \circ g) = 0 = 0_Y(\beta) \text{ for each } \beta \in 2^Y.$$

Thus $0_Y = Gg(0_{2^2}) \in Gg[A] \subset GY$.

M. Fine, boy! Now let us see that F has the properties (A1) – (A4).

S. I love to, Masta! Obviously $|F0| = 1$. Next I will show:

(1) $\quad F2 = A$.

Consider the maps $f_{ij}\colon 2 \longrightarrow 2$, defined by $f_{ij}(0) = i$ and $f_{ij}(1) = j$, for $(i,j) \in 2^2$. Then a straightforward computation shows:

---

[2]For notational convenience, elements f of 2^2 are identified with the corresponding elements $(f(0), f(1))$ of 2×2. Thus, e.g., $(0,0)$ denotes the constant map $f\colon 2 \longrightarrow 2$ with value 0; etc.

$$\begin{array}{lcll} Gf_{ij}(0_{2^2}) & = & 0_{2^2} & \text{for each pair } (i,j), \\ Gf_{ij}(a) & = & 0_{2^2} & \text{for } i = j, \\ Gf_{ij}(a) & = & a & \text{for } i \neq j. \end{array}$$

Thus $F2 = A$.

Next I will show:

(2) $\quad F1 = \{0_{2^1}\}$.

Let $f\colon 2 \longrightarrow 1$ be the unique map from 2 to 1. Then $f(0) = f(1) = 0$. Thus

$$\big(Gf(a)\big)(\beta) = a(\beta \circ f) = a\big(\beta(f(0)),\ \beta(f(1))\big) = \beta\big(f(0)\big) + \beta\big(f(1)\big) =$$

$$= \beta(0) + \beta(0) = 0 \pmod 2 \text{ for each } \beta \in 2^1.$$

Thus $Gf(a) = 0_{2^1}$.

Further: $\big(Gf(0_{2^2})\big)(\beta) = 0_{2^2}(\beta \circ f) = 0 = 0_{2^1}(\beta)$ for each $\beta \in 2^1$.

Next I will show:

(3) $\quad |Fn| = \binom{n}{2} + 1$ for $2 < n < \infty$.

Consider the n^2 maps $g_{ij}\colon 2 \longrightarrow n$, defined by $g_{ij}(0) = i$ and $g_{ij}(1) = j$, where $(i,j) \in n^2$.

Then:

$$\big(Gg_{ij}(0_{2^2})\big)(\beta) = 0_{2^2}(\beta \circ g_{ij}) = 0 = 0_{2^n}(\beta) \text{ for each } \beta \in 2^n.$$

Thus $Gg_{ij}(0_{2^2}) = 0_{2^n}$ for each pair (i,j).

$$\big(Gg_{ij}(a)\big)(\beta) = a(\beta \circ g_{ij}) = \beta(i) + \beta(j) (\mathrm{mod}\ 2).$$

Thus:

α) $Gg_{ij}(a) = 0_{2^n}$ for $i = j$,

β) $Gg_{ij}(a) = Gg_{ji}(a) \neq 0_{2^n}$ for $i \neq j$,

γ) $Gg_{ij}(a) \neq Gg_{kl}(a)$ for $i \neq j,\ k \neq l$, and $\{i,j\} \neq \{k,l\}$.

Consequently Fn has precisely $1 + \binom{n}{2}$ elements.
Next I will show:

(4) $\quad |F| = |X| \qquad$ for infinite X.

As above we conclude that, if X is infinite and $\mathcal{P}_2 X$ is the set of all 2–element subsets of X, then $|FX| = 1 + |\mathcal{P}_2 X| = 1 + |X|^2 = 1 + |X| = |X|$. See, Masta, now all is shown.

M. Indeed, you have constructed an endofunctor F of **Set** that satisfies the conditions (A1) – (A4) formulated above, — and thus you have provided a positive answer to my problem 5a. In fact, you have done more. Your functor F can be considered as a *minimal* endofunctor of **Set** that satisfies conditions (A1) and (A2), since Václav Koubek[3] has shown long ago that any F which satisfies (A1) and (A2) must also satisfy:

(A3)* $|Fn| \geq \binom{n}{2} + 1$ for each n = with $2 < n < \infty$
[Lemmata 1.2 and 3.3]

(A4)* $|FX| \geq |X|$ for infinite X
[Lemma 4.3]

Moreover, V. Koubek has demonstrated (Theorem 1.3) that (A3) and (A4) follow from (A1), (A2) and the additional assumption (A0):

(A0) $FX = \cup\{f[Y] \mid Y \xrightarrow{f} X \text{ a map with } \lvert Y\rvert < \lvert X\rvert\}$ for each set X with $2 < \lvert X\rvert$.

So we may ask:

Problem 6: Let F be an endofunctor of **Set**.
a) Is F uniquely[4] determined by the properties (A1) – (A4)?
b) Is F uniquely[3] determined by the properties (A1) and (A2)?
c) Is F uniquely[3] determined by the properties (A0) – (A2)?

Next day

S. Master — the answer to your problem must be no! Last night my buddy Walter pointed out to me an entirely different functor F that satisfies the conditions (A1) – (A4). In fact, his construction is much simpler: Let FX be the disjoint union of the set $\{0\}$ and the set of all 2–element subsets of X; and for any map $f\colon X \longrightarrow Y$ define

$$Ff(B) = \begin{cases} f[B], & \text{if } |B| = |f[B]| = 2, \\ 0, & \text{otherwise.} \end{cases}$$

Then F satisfies (A1) – (A4).

M. A very elegant construction indeed — and quite different from yours which, by the way, — was known to sweet Věra for a long time. But are you sure that your F and Walter's F are not just different but *categorically* different? Couldn't it be, e.g., that they are naturally equivalent?

---

[3] V. Koubek, *Set Functors.* Comment. Math. Univ. Carolinae **12** (1971) 175–195

[4] *Uniqueness* to be understood in categorical sense as *essential uniqueness.*

An hour later

S. Master! Master! — I am so sorry! You are absolutely right! Walter's and my functors are naturally equivalent. Call Walter's functor W and my functor F, then the bijections

$$\eta_X \colon WX \longrightarrow FX$$

defined by

$$\eta_X(B) = \begin{cases} 0_{2^X}, & \text{if } \quad B \quad = \quad 0, \\ Gf_B^X(a), & \text{if } \quad |B| \quad = \quad 2, \end{cases}$$

where $f_B^X \colon 2 \longrightarrow X$ is one of the two maps from 2 to X with image B (it doesn't matter which, since $Gf_B^X(a)$ doesn't depend on the choice), form a natural equivalence between W and F. Do you want to see the proof?

M. Perhaps later, son. I trust that you checked things carefully enough. I would rather like to know the answer to Problem 6.

However, let us first return to Problem 5. Since your answer to Problem 5a is positive, so is the one to Problem 5c. This leaves Problem 5b. Did you solve that, too?

S. Oh yes— Masta! Honky–tonk stuff after solving 5a! In fact I have constructed another endofunctor F of **Set** with the following properties:

(B1) $\lvert F0\rvert = \lvert F1\rvert = [F2] = 1$,
(B2) $\lvert F3\rvert = 3$,
(B3) $\lvert Fn\rvert = 1 + 2\binom{n}{3}$ for each n with $3 < n < \infty$,
(B4) $\lvert FX\rvert = \lvert X\rvert$ for infinite X,

The construction proceeds exactly as before with the following two modifications:

1. The role of 2 in the construction of G is taken over by 3.
2. In the construction of the subfunctor F of G, condition (b) is replaced by:

$$\boxed{\begin{array}{l} FX = \cup\{Ff[A] \mid f \in 3^X\} \text{ for } X \neq \emptyset, \text{ where } A = \{a_0, a_1, a_2\} \text{ with} \\ a_i \colon 3^3 \longrightarrow 3 \text{ being defined by } a_i(\alpha) = \begin{cases} i, & \text{if } \alpha \text{ is surjective,} \\ 0, & \text{otherwise.} \end{cases} \end{array}}$$

This, Masta, provides a complete answer to your Problem 5.

M. Well done, boy! You may start to compete with Věra now. She has sent me a different solution for Problem 5b. However your answers to Problems 4 and 5 raise another obvious question. Do you see it?

S. Well Masta, — I think so. I have shown that

(a) there is no functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ with $|F1| = 1$, $|F2| = 2$, and $|F3| = 3$,

but that there are functors $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ with resp.

(b) $|F1| = 1$ and $|F2| = 2$,
(c) $|F1| = 1$ and $|F3| = 3$.

So the following unnatural problem forces itself on me:

> **Problem 7:** Does there exist a functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Inj}$ with $|F2| = 2$ and $|F3| = 3$?

M. Very good, boy! You have learned to create problems by yourself. — Did you solve the above already?

S. Well, Masta, my friend Evangeliste, after seeing my construction of the functor F which solves your Problem 5b, was able to modify it as follows: if a_2 is replaced by the constant function with value 1, then the resulting functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Set}$ satisfies the conditions:

> (C1) $|F0| = 1$ and $|F1| = 2$,
>
> (C2) $|F2| = 2$ and $|F3| = 3$,
>
> (C3) $|Fn| = 2 + \binom{n}{3}$ for each n with $3 < n < \infty$,
>
> (C4) $|FX| = X$ for each infinite set X.

M. Quite clever, your friend. I never heard of him before.

S. Masta — you must be getting old. He is a close friend of Dave and Dexue and helped you to discover the famous result that $\mathbb{N}$ is Lindelöf if and only if the classical Ascoli Theorem holds.

M. Oh, sorry. Yes, I remember now. I remember him well, indeed! Evangeliste Jimplecute, this teeny–bopper heartthrob, – a fine fellow indeed. But what a strange name. I recall reading that in the Ozarks "there are men still living who remember stories of the jimplecute, a kind of ghostly dragon or dinosaur supposed to walk the roads at night, grab travelers by the throat and suck their blood. Some say this creature was invented near Argenta, Arkansas, in the 1870's, to frighten superstitious Negroes."[5]

---

[5]From *AR Hist. Quart. 9.69 Ozarks.*

The xanthus joshed the jimplecute,
The aardvark hugged the auk,
The woozle woggled in and out,
To snare the stubenrauch.[6]

S. Masta, oh Masta! Don't get carried away – please!
Have we exhausted all problems now or are there more?

M. Well boy, it seems that you don't ever get hardly able fever.

S. Masta, what is that?

M. Don't you know? It is the illness of feeling lackadaisical.

As far as problems are concerned, they never get exhausted. You may ask, e.g.:

Problem 8: For which classes $\mathcal{C}$ of cardinals does there exist an endofunctor F of **Set** with

$$\mathcal{C} = \{|X| \mid |FX| = |X|\}?$$

Problem 9: Are the solutions of Problem 8, perhaps under some natural restrictions, essentially unique?

Problem 10: Replace in our problems the category **Set** by some other category, say **Pos**. What happens then? Are the results similar? If so, which properties of the categories in question are responsible for this?

So you see, son, there is no end to the creation and evolution of ideas. Keep going, boy! Keep going! You are on your own feet now. But if you need help, ask Věra[7], the masterful constructor of set–functors, weird products, and similar things who requires her students to construct an endofunctor of Set with $F(13) = 11$ and $F(17) = 31$, or get in contact with Alois[8] who knows quite as much as Věra about these — and other — matters.

S. Masta, one more thing: I showed my notes also to my friend Dibbler T. Referee. He looked them over and here is the result. Didn't he do a wonderful job?

[6] From *Sat. Review* Oct. 2, 1926.

[7] Cf. V. Trnková, *Some properties of Set–functors.* Comment.Math. Univ. Carolinae **10** (1969) 323–352.
—, *A descriptive classification of Set–functors I, II.* Comment. Math. Univ. Carolinae **12** (1971) 143–174 and 345–357.

[8] Alois Zmrzlina, *Too Many Functors — A Continuation of "The Emergence of Functors".* This volume.

M. Perfect, my son. Just perfect!

> "Moral: It is better to ask some of the questions than to know all the answers."
>
> From *The Scotty who knew too much*
> by James Thurber

Y.T. Rhineghost
c/o H. Herrlich
Department of Mathematics
University of Bremen
P.O. Box 33 o4 4o
28334 Bremen, Germany
herrlich@math.Uni-Bremen.de

Too Many Functors

— A Continuation of "The Emergence of Functors" —

Alois Zmrzlina

Dedicated to George E. Strecker — I wish him good health, many good students and a mountain of good ice cream.

Key words: Endofunctor of SET, category of sets, fixpoint, injective hull.

AMS Subject Classifications (2000): 18A22, 18B05.

S. Masta, in your last lecture you gave us the references [1] – [7] about set–functors **Set** $\longrightarrow$ **Set**. But it is too much to read them all. I looked at [5] and [6] because these are the oldest ones, so they could be the beginning of the story. I've seen some more or less interesting functors there. But do you know, Masta, what is my favorite set–functor?

M. Yes?

S. It is the constant to the empty set C_0.

M. What is so interesting about this trivial functor?

S. It forms so many exceptions that you must never forget it.

M. Really?

S. It is the unique functor which sends a non–empty set to $\emptyset$. All the other functors admit the domain–range restriction to **Inj** $\longrightarrow$ **Inj**.

M. You are right. What else?

S. It is the unique functor which is naturally equivalent only to itself.

M. That's true. What else?

S. It is the unique functor F which does not permit a transformation of the identity functor Id into F.

M. It really does not permit such a transformation. But why is it the unique one?

S. For all the other functors F, the set $F(1)$ is non–empty. Let $v_0, v_1\colon 1 \longrightarrow 2$ denote the maps which send the unique element of $1 = \{0\}$ to 0 or 1, where $2 = \{0,1\}$. If $F(v_0) = F(v_1)$, then there is a transformation $C_{0,1} \longrightarrow F$, by [6, II.1, II.3], where $C_{0,1}$ is the functor sending $\emptyset$ to itself and all the other sets to 1. And $C_{0,1}$ is, clearly, a factor-functor of Id (i.e., there is an epitransformation $\mathsf{Id} \to C_{0,1}$). And if $F(v_0) \neq F(v_1)$, then there is even a monotransformation $\mu\colon \mathsf{Id} \longrightarrow F$. In fact, we choose $a \in F(1)$ with $[F(v_0)](a) \neq [F(v_1)](a)$. For every set X with $|X| \geq 2$ and $x_0, x_1 \in X$, $x_0 \neq x_1$, we put $v_{x_0} = h \cdot v_0$ and $v_{x_1} = h \cdot v_1$ where $h\colon 2 \longrightarrow X$ sends 0 to x_0 and 1 to x_1. Since h is a coretraction in **Set**, $F(h)$ must be one–to–one, so that $[F(v_{x_0})](a) \neq [F(v_{x_1})](a)$. Hence, $\mu_X\colon X \longrightarrow F(X)$ sending each $x \in X$ to $[F(v_x)](a)$ determines the monotransformation μ.

M. Very good, my boy. We summarize it in a useful statement:

> If $F \neq C_0$, then either F contains an internal copy of $C_{0,1}$ or an internal copy of Id,

where the internal copy of $C_{0,1}$ or Id means that F has a subfunctor naturally equivalent to $C_{0,1}$ or Id.

S. Masta, Masta, I still haven't finished the list of the remarkable properties of the functor C_0. It is the unique functor which preserves limits of non–empty diagrams but not all limits.

M. Really? Tell it, please.

S. Clearly, C_0 preserves limits of non–empty diagrams and it does not preserve the one–point set 1, i.e., the terminal object of **Set** which is a limit of the empty diagram. If F is a functor distinct from C_0, then necessarily $F(1) \neq \emptyset$. Choose a system $\{1_i \mid i \in I\}$ of $|I|$ copies of 1 with $|I| > |F(1)|$, so that $1 = |\prod_{i\in I} 1_i|$. If F preserves non–empty products, then necessarily $|F(1)| = \prod_{i\in I} |F(1_i)|$. This is possible only when $|F(1)| = 1$. Thus F preserves the terminal object, i.e., it preserves the limit of the empty diagram as well, so that if F preserves limits of all non–empty diagrams, it preserves all limits.

M. Very good, C_0 is really an exceptional functor.

S. Masta, I think that the functor $C_{0,1}$ is also nice. It is the unique functor $\mathbf{Set} \longrightarrow \mathbf{Set}$ which preserves all products but not all limits. It does not preserve the empty equalizers.

M. That's quite easy that $C_{0,1}$ preserves all products but not all limits. But why is it the unique one? You surely mean the uniqueness up to natural equivalence.

S. Certainly, Masta. The uniqueness is proved in [5, Theorem 4.1].

M. Limits can be constructed from products and equalizers, by the well-known Maranda construction. Hence, a functor from a complete category preserves all limits if and only if it preserves all products and all equalizers. Hence, the preservation of all equalizers is, in a way, a "complementary" property to the preservation of all products. Is a set–functor preserving all equalizers but not all limits also a unique one?

S. No, Masta. The functors preserving equalizers are characterized in [6] and there are plenty of set–functors which preserve all equalizers but not all limits. For example, if $\mathcal{F}$ is a free filter on an infinite set M, then the factor-functor $Q_{M,\mathcal{F}}$ of the hom–functor $\hom(M,-)$, obtained by gluing of any pair of maps $f, g\colon M \longrightarrow X$ which coincide on a set of $\mathcal{F}$, preserves all equalizers but not all products. Since there are so many such functors, they don't interest me too much.

M. But I don't think that only exceptions are interesting. What do you think about the Koubek's functors B_α^β?

S. I've never heard about them. Tell me, Masta, please!

M. Let us denote by P the power–set functor, i.e., for every set X, $P(X)$ is the set $\exp X$ of all subsets of X and, for every map $f\colon X \longrightarrow Y$ and every $Z \in P(X)$, $[P(f)](Z)$ is the image $f(Z)$. For every cardinal $\alpha > 0$, let $P_{\leq\alpha}$ and $P_{<\alpha}$ denote the subfunctors of P such that, for every X,

$$P_{\leq\alpha}(X) = \{Z \subseteq X \mid |Z| \leq \alpha\} \text{ and } P_{<\alpha}(X) = \{Z \subseteq X \mid |Z| < \alpha\}.$$

If $\alpha \leq \beta$, then B_α^β is the factor-functor of $P_{\leq\beta}$ obtained from it by the collapsing of its subfunctor $P_{<\alpha}$ into C_1 where C_1 is the constant functor sending every set to 1. You should think about **these** functors and you should read the paper [1].

S. O.K., Masta, I'll do it.

Next day

S. Masta, you were right. The functors B_α^β are extremely interesting, especially the case $\alpha = \beta$. I'll denote B_α^α simply by B_α and call the copy of C_1 obtained by the collapsing of $P_{<\alpha}$ the internal copy of C_1 in B_α. The functors B_α give simple solutions of several of our previous problems: B_2 solves positively Problem 5a in "The Functor that Wouldn't be", two copies of B_3 glued together along their internal copies of C_1 solve positively Problem 5b and a coproduct of C_1 and B_3 solves positively Problem 7 in "The Emergence of Functors". Moreover, they lead immediately to the solution of Problem 6b; there are plenty of functors satisfying the conditions (A1) and (A2) in "The Emergence of Functors", e.g., B_2 glued along C_1 with B_α where $\alpha > 2$ or with many such B_α's.

M. So I see that you worked a lot. And what about functors satisfying (A1) – (A3)?

S. There are also many functors satisfying (A1) – (A3). The functor F obtained from gluing of B_2 with B_α along their internal copies of C_1 satisfies (A1) – (A3) whenever α is infinite. Since $|F(\alpha)| \geq |B_\alpha(\alpha)| > \alpha = |B_2(\alpha)|$ the functors F and B_2 are not naturally equivalent.

M. And if a functor satisfies (A1) – (A4), is it already uniquely determined?

S. Exactly. If F is a functor with $|F(\emptyset)| = 1 = |F(1)|$, then F has precisely one internal copy of C_1. Since $|F(2)| = 2$, there is precisely one element of $F(2)$ which is not in the internal copy C_1. Let me denote it by a. You can define a natural transformation

$$\tau \colon B_2 \longrightarrow F$$

such that it sends the internal copy of C_1 in B_2 onto the internal copy of C_1 in F and its 2nd component $\tau_2 \colon B_2(2) \longrightarrow F(2)$ sends the element $2 = \{0, 1\} \in B_2(2)$ to a. Hence, for every set X with $|X| \geq 2$, the X–th component τ_X of τ is defined by

$$\tau_X(\{x_0, x_1\}) = [F(v_{x_0,x_1})](a)$$

for any pair of distinct elements $x_0, x_1 \in X$ where $v_{x_0,x_1} \colon 2 \longrightarrow X$ sends 0 to x_0 and 1 to x_1. This definition is correct because if $f, g \colon 2 \longrightarrow X$ are maps with

$$[B_2(f)](2) = [B_2(g)](2),$$

then necessarily

$$[F(f)](a) = [F(g)](a).$$

This is because the former equation implies that $f = g \circ \sigma$, where $\sigma \colon 2 \longrightarrow 2$ is either the identity (if $f = g$) or the isomorphism exchanging 0 and 1 (if $f \neq g$). Since $|F2| = 2$ and σ is an isomorphism, necessarily $[F(\sigma)](a) = a$. This implies the latter equation. Hence τ is really a natural transformation.

M. That's right. Continue, please!

S. For all finite X, τ_X must be one–to–one, otherwise the image $\tau(B_2)$ would be a subfunctor of F not satisfying (A3)* of "The Emergence of Functors". But τ_X must be one–to–one also for infinite X's because if $\tau_X : B_2(X) \longrightarrow F(X)$ glues $\{x_0, x_1\} \in B_2(X)$, for some $x_0 \neq x_1$, with some $\{y_0, y_1)\} \in B_2(X)$ (where not necessarily $y_0 \neq y_1$, i.e., $\{y_0, y_1\}$ can be in the internal copy of C_1 in B_2) and if $\{x_0, x_1\} \neq \{y_0, y_1\}$, we would take $f : 4 \longrightarrow X$ (or $3 \longrightarrow X$ or $2 \longrightarrow X$ depending on the cardinality of $\{x_0, x_1, y_0, y_1\}$) which is one–to–one and sends 4 (or 3 or 2) onto $\{x_0, x_1, y_0, y_1\}$. Then $F(f)$ has to be also one–to–one and $F(f) \circ \tau_4 = \tau_X \circ B_2(f)$ (and analogously, for τ_3 or τ_2). Since τ_4 (or τ_3 or τ_2) is one–to–one, τ_X cannot glue $\{x_0, x_1\}$ with $\{y_0, y_1\}$. Hence, if F satisfies (A1) and (A2), $\tau : B_2 \longrightarrow F$ is a monotransformation.

M. So, you proved in fact that

if a functor satisfies (A1) and (A2), it contains an internal copy of B_2.

S. Yes. And if F satisfies also (A3), the X–th components $\tau_X : B_2(X) \longrightarrow F(X)$ of the above monotransformation $\tau : B_2 \longrightarrow F$ must be surjective for all finite X.

M. That's clear. But what about the infinite sets X?

S. We use Koubek's notion of unattainable cardinal. For any set–functor H and any set X, Koubek in [1] investigates the set

$$H(X) \setminus \cup[H(f)](H(Y)),$$

where the union is over all maps $f : Y \longrightarrow X$ with $|Y| < |X|$ and he calls this set an *increase* of H at X. And a cardinal $|X| > 1$ is called *unattainable* for H iff the increase of H at X is non–empty. Hence, e.g., B_α with $\alpha \geq 2$ has precisely one unattainable cardinal, namely α. And the condition (A0) of "The Emergence of Functors" is precisely the requirement that no $\alpha > 2$ is an unattainable cardinal of H.

M. Well, and how do you use this?

S. Koubek proved in [1] the following useful statement:

$$(US)\left\{\begin{array}{l}\text{If } \alpha \geq 2 \text{ is an unattainable cardinal of a functor } H \text{ and} \\ a \text{ is an element of the increase of } H \text{ at } \alpha, \text{ then, for every} \\ \text{set } X \text{ and every two one-to-one maps } f, g\colon \alpha \longrightarrow X \\ \text{such that the intersection of their images is small, i.e.,} \\ |\mathrm{Im} f \cap \mathrm{Im} g| < \alpha, [H(f)](a) \text{ and } [H(g)](a) \\ \text{are distinct elements of } H(X).\end{array}\right.$$

Hence, for an infinite unattainable cardinal $|X|$, we get

$$|H(X)| > |X|$$

because the above statement, with $\alpha = |X|$, gives a one–to–one map of every almost disjoint system $\mathcal{X}$ on X (i.e., a system $\mathcal{X} \subseteq \exp X$ such that $|Y| = |X|$ for every $Y \in \mathcal{X}$ and $|Y_1 \cap Y_2| < |X|$ for every distinct $Y_1, Y_2 \in \mathcal{X}$) into $H(X)$. And it is well known that, for every infinite set X, there is an almost disjoint system $\mathcal{X}$ on X with $|\mathcal{X}| > |X|$.

M. This is really a very useful statement.

S. Yes, Masta. If F satisfies (A1) – (A4), then the above monotransformation $\tau\colon B_2 \longrightarrow F$, with $\tau_X\colon B_2(X) \longrightarrow F(X)$ surjective for all finite sets X, has its components τ_X surjective also for infinite sets X; otherwise F would have an infinite unattainable cardinal in contradiction with (A4).

M. Thus, you solved all the three questions of Problem 6. Very good, my boy, you made a very good job. Now, you can have a look at systems of functorial equations.

S. System of functorial equations? Masta, what is it?

M. It is a system of the form

$$\begin{array}{rcl}\mathbb{F}(n_0) & = & m_0, \\ \mathbb{F}(n_1) & = & m_1, \\ & \vdots & \end{array}$$

where n_i, m_i are cardinal numbers. A set–functor F is its solution if $|F(n_i)| = m_i$ for all the equations of the system. And the system is solvable iff it has a solution.

S. And you mean that I could investigate whether a system is solvable or not. I've already solved the system

$$\begin{array}{rcl}\mathbb{F}(0) & = & 1, \\ \mathbb{F}(1) & = & 1, \\ \mathbb{F}(2) & = & 2;\end{array}$$

its solution is B_2, but if I add the equation

$$\mathbb{F}(3) = 3,$$

the obtained system is no longer solvable, as I proved in "The Functor that Wouldn't be".

M. That's right. But you can investigate more general situations, for example the system

$$\begin{aligned} \mathbb{F}(8) &= 5, \\ \mathbb{F}(9) &= 7. \end{aligned}$$

S. Well, ... (*after 10 minutes*) .... This system is solvable, Masta! The coproduct of two copies of B_9 and three copies of C_1 forms a solution!

M. And what about system

$$\begin{aligned} \mathbb{F}(8) &= 7, \\ \mathbb{F}(9) &= 5? \end{aligned}$$

S. This system is not solvable because the embedding $f\colon 8 \longrightarrow 9$ is a co-retraction hence, for a possible solution F, $F(f)\colon 7 \longrightarrow 5$ has to be one–to–one which is impossible.

M. This was easy, wasn't it? But you could have a look at

> **Problem 11.**
> Depending on a given quadruple (n_0, n_1, m_0, m_1) of cardinal numbers, discuss the solvability of the system
>
> $$\begin{aligned} \mathbb{F}(n_0) &= m_0, \\ \mathbb{F}(n_1) &= m_1! \end{aligned}$$

We may suppose $n_0 \leq n_1$, of course.

S. O.K., Masta, I'll try it.

After 3 days

S. Masta, I've almost solved Problem 11.

M. "Almost solved" — what does it mean?

S. Let me start to discuss the solvability of the system of Problem 11, please. If $n_0 = n_1$, then, clearly, the system is solvable iff $m_0 = m_1$. Hence we may suppose $n_0 < n_1$. If $n_0 = 0$, then the system is solvable unless $m_0 > 0 = m_1$: if $m_0 = m_1 = 0$, then C_0 solves it; if $m_1 \neq 0$, then the functor sending $\emptyset$ to m_0 and all non–empty sets to m_1 (and the empty maps $\emptyset \longrightarrow X \neq \emptyset$ to an arbitrary fixed map $p\colon m_0 \longrightarrow m_1$) solves it.

M. That's quite clear. Hence you may suppose $0 < n_0 < n_1$.

S. Exactly. If $m_0 > m_1$, the system is not solvable because the image of the inclusion $n_0 \longrightarrow n_1$ under any functor must be a one–to–one map. If $m_0 = m_1$, then the system **is** solvable, the constant functor to m_0 solves it. If $m_0 = 0 < m_1$, then the system is not solvable because C_0 is the unique functor sending $n_0 > 0$ to 0 and this contradicts the condition $\mathbb{F}(n_1) = m_1$.

M. Thus, you can suppose $0 < n_0 < n_1$ and $0 < m_0 < m_1$.

S. Yes, non–trivial cases start now. For every pair of cardinal numbers $p \geq 2$ and $q \geq 1$, let us denote by C_p the coproduct of p copies of the functor C_1 and by $B[p, q]$ the functor obtained from q copies of B_p by gluing them all together along their internal copy of C_1 (hence $B[p, 1]$ is just B_p). Given $0 < n_0 < n_1$ and $0 < m_0 < m_1$ and n_1 is finite, then the system is solvable, the coproduct of $B[n_1, m_1 - m_0]$ and $C_{m_0 - 1}$ solves it.

M. This case includes the system

$$\begin{aligned} \mathbb{F}(13) &= 11, \\ \mathbb{F}(17) &= 31 \end{aligned}$$

mentioned in "The Emergence of Functors".

S. Now, let us suppose that $0 < n_0 < n_1$, $0 < m_0 < m_1$ and, moreover, n_1 is infinite. If $m_1 \geq 2^{n_1}$, then the system is solvable, the coproduct of $B[n_1, m_1]$ and $C_{m_0 - 1}$ solves it. For the case $m_1 < n_1$, I proved the following useful statement:

If H is a functor with $\lvert H(n) \rvert < n$ for some infinite n, then the restriction of H to the category $\mathbf{Set}_{\leq n}$ of all sets of cardinality less than or equal to n is constant.

Should I show you the proof?

M. No, my boy. V. Koubek in [1] already proved it. You overlooked it there.

S. This statement implies immediately that, for $m_1 < n_1$, the system is not solvable because we suppose $m_0 < m_1$.

M. That's right. Hence, if you suppose that there is no cardinal between n_1 and 2^{n_1}, only the case $n_1 = m_1$ is missing.

S. This assumption is a part of my "almost", Masta. For the case $0 < n_0 < n_1$, $0 < m_0 < m_1$ and n_1 infinite, I really suppose the Generalized Continuum Hypothesis. Thus, under (GCH), let us go to the remaining most complicated case $n_1 = m_1$. If n_0 is finite, then the system is solvable, the functors $B[n_0, m_0 - 1]$ for $m_0 > 1$ and B_{n_0+1} for $m_0 = 1$ solve it.

M. Hence only the case $\aleph_0 \leq n_0 < n_1 = m_1 > m_0 > 0$ is missing.

S. If $m_0 \geq n_0$, then the system is solvable, the coproduct of m_0 copies of the identity functor Id solves it. In the remaining case, $m_0 < n_0$. Let us denote by n_0^+ the successor of n_0. If $n_1 = n_0^+$, then the system is not solvable. In fact, a possible solution F has to be constant on $\mathbf{Set}_{\leq n_0}$. Since $m_0 < m_1$, n_1 has to be an unattainable cardinal of F, $|F(n_1)| > n_1$, which is a contradiction. If $n_0^+ < cf\ n_1$, where $cf\ n_1$ denotes the cofinality of n_1, the system is solvable and the coproduct of $B_{n_0^+}$ and C_{m_0-1} solves it. However, the case $cf\ n_1 \leq n_0^+ < n_1$ *remains open.* For example, I am not able to decide the solvability of the system

$$\begin{aligned} \mathbb{F}(\aleph_1) &= \aleph_0, \\ \mathbb{F}(\aleph_\omega) &= \aleph_\omega. \end{aligned}$$

M. But, as I see, your discussion is complete if you restrict to regular cardinals n_1.

S. Yes, Masta. Under (GCH), a machine can solve whether, for a given quadruple (n_0, n_1, m_0, m_1) with n_1 finite or regular, the system

$$\begin{aligned} \mathbb{F}(n_0) &= m_0, \\ \mathbb{F}(n_1) &= m_1 \end{aligned}$$

is solvable or not, even not knowing the values n_0, m_0, n_1, m_1 precisely. The machine only asks you several questions about the quadruple, at most eleven. You can answer only by pressing a YES — NO button, and then the machine tells you whether the system *is* or *is not* solvable. And if it is solvable, the machine tells you a functor solving the system.

But my machine gives results also for many quadruples with n_1 being a singular cardinal number. It says "I do not know" *only* in case

$$0 < m_0 < n_0, \ cf\ n_1 \leq n_0^+ < n_1 \text{ and } m_1 = n_1.$$

M. You did a very good job, my boy. Are you not yet tired from the solving of systems of functorial equations?

S. No, Masta. I would like also to have a look at the following

> **Problem 12.**
> Discuss the solvability of systems of three (or more) functorial equations!

I'd like to look at it at least for n_i's finite.

M. Though restricted to finite cardinals, it could be a pretty complicated problem. But there are other interesting problems concerning systems of functorial equations. We say that such a system is *uniquely solvable* iff it has precisely one solution, one up to natural equivalence, of course. You can have a look at the following

> **Problem 13.**
> Find a necessary and sufficient condition for the unique solvability of a system consisting of a *set* of functorial equations $\mathbb{F}(n_i) = m_i$!

Next day

S. Masta, I fully solved Problem 13. Any system consisting of a **set** of functorial equations is never uniquely solvable unless my favorite C_0 forms its unique solution, i.e., unless all the m_i's are equal to zero while $n_i \neq 0$ for at least one n_i!

M. In this case, such a system is really uniquely solvable and C_0 is its unique solution.

S. Yes, and, in all other cases, either the system $\mathbb{F}(n_i) = m_i$, with i ranging over a set I, is not solvable at all or a functor F distinct from C_0 solves it. In the latter case, I'll find a functor G which also solves the system and G is not naturally equivalent to F. Choose cardinal numbers α with $\alpha > n_i$ for all $i \in I$ and γ with $\gamma > |F(\alpha)|$. Denote by H the functor such that

$$H(\emptyset) = \emptyset \text{ and } H(X) = B[\alpha, \gamma](X) \text{ for all } X \neq \emptyset.$$

Hence H contains a unique internal copy of $C_{0,1}$. Since $F \neq C_0$, F contains either an internal copy of $C_{0,1}$ or an internal copy of Id. In the former case, we glue F and H along their internal copies of $C_{0,1}$; in the latter case, we glue F and $H \times \mathsf{Id}$ along their internal copies of Id. In both cases, the obtained functor G also solves the system because α is very large. And, since $\gamma > |F(\alpha)|$, the functor G is not naturally equivalent to F because $|G(\alpha)| > |F(\alpha)|$.

M. Very good, my boy. So only large systems of functorial equations are of interest from the point of view of the unique solvability. The main problem concerning the unique solvability is as follows: every functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Set}$ determines a large solvable system, namely

$$\mathbb{F}(n) = |F(n)| \text{ for all cardinal numbers } n.$$

Let us call it the system of F. Has this system also a solution other than F? It is meaningful to expect that, for some functors, the answer is YES and, for some functors, the answer is NO.

S. Masta, Masta, I see a functor with YES and a functor with NO!

M. Really? Tell me, please!

S. I proved that (A1) – (A4) in "The Emergence of Functors", which form the system of functorial equations of B_2, is uniquely solvable. But the system of the coproduct of two copies of B_2 is not uniquely solvable. It can be solved also by the coproduct of C_1 with the functor $B[2,2]$.

M. Very good.

S. I guess that the system of a coproduct of two copies of a non–constant functor is not uniquely solvable.

M. And what about the coproduct of two copies of the identity functor Id?

S. Well, ... it really looks like its system is uniquely solvable. I'll think about it.

M. First, you should think about the unique solvability of the system of the functor Id itself.

S. That's right. The system of Id is surely uniquely solvable. Well, ... how to prove it? .... I see it, it's easy! Let F be a functor with $|F(n)| = n$ for all cardinal numbers n. Since $F \neq C_0$, it contains either an internal copy of $C_{0,1}$ or of Id. Since $|F(1)| = 1$, F contains precisely one such copy. In the former case, we can redefine F at $\emptyset$ putting $F(\emptyset) = 1$ (we denote the redefined functor by F again). Then $F(\emptyset) = 1 = |F(1)|$, $|F(2)| = 2$ and $|F(3)| = 3$. This is impossible as I proved in "The Functors that Wouldn't be". In the latter case, we have a monotransformation $\mu\colon \mathsf{Id} \longrightarrow F$. For all finite X, the Xth component $\mu_X\colon X \longrightarrow F(X)$ must be surjective because $|F(X)| = |X|$. For infinite X, we use Koubek's result about unattainable cardinals again: if α is the smallest infinite cardinal for which $\mu_\alpha\colon \alpha \longrightarrow F(\alpha)$ is not surjective, then α is an unattainable cardinal of F hence $|F(\alpha)| > \alpha$ which is a contradiction. Hence μ is a natural equivalence of F and Id.

M. Very good. Now, let's go back to the functor $\mathsf{Id} + \mathsf{Id}$.

S. Well, let G be a functor with $|G(n)| = 2n$ for all n. We can express G, as in [6], in the form $G_0 + G_1$ with $|G_0(1)| = |G_1(1)| = 1$ (by means of the preimages of 0 and 1 in $|G(1)| = 2$ under the maps $G(f)$ with $f\colon X \longrightarrow 1$, for every X). Then $|G_0(2)| + |G_1(2)| = |G(2)| = 4$. If $|G_0(2)| = 3$ we can use the fact, proved in [5], that every set–functor preserves finite non–empty intersections. Hence lifting by G_0 the diagram

$$\begin{tikzcd}
\{0\} \arrow[r] \arrow[dr] & \{0,1\} \arrow[dr] & \\
\{1\} \arrow[ur] \arrow[dr] & \{0,2\} \arrow[r] & \{0,1,2\} = 3 \\
\{2\} \arrow[r] \arrow[ur] & \{1,2\} \arrow[ur] &
\end{tikzcd}$$

consisting of the inclusions, we get that $|G_0(3)| \geq 6$. Since $|G_1(3)| \geq 1$, we get $|G(3)| \geq 7$ which is a contradiction. Analogously, if $|G_1(2)| = 3$. Hence necessarily $|G_0(2)| = |G_1(2)| = 2$.

M. That's right. Continue, please!

S. Let us suppose that G_0 contains an internal copy of $C_{0,1}$. Hence we can redefine G_0 at $\emptyset$ putting $G_0(\emptyset) = 1$ (we denote this redefined functor by G_0 again) hence G_0 satisfies (A1) and (A2) of "The Emergence of Functors", hence, by (A3)*, $|G_0(n)| \geq \binom{n}{2} + 1$ so that $|G(n)| > 2n$ for all finite $n > 2$ which is a contradiction. Analogously we get a contradiction if G_1 contains an internal copy of $C_{0,1}$. In the remaining case, we have monotransformations $\mu^0\colon \mathsf{Id} \longrightarrow G_0$ and $\mu^1 : \mathsf{Id} \longrightarrow G_1$. The proof that all components $\mu^0_X\colon X \longrightarrow G_0(X)$ and $\mu^1_X : X \longrightarrow G_1(X)$ have to be surjective is quite analogous to the above case, i.e., for Id only.

M. Very good, my boy. Thus, you proved in fact that the systems of the functors $T_M(-) = - \times M$ are uniquely solvable whenever $|M| = 1$ or $|M| = 2$; for $|M| = 0$ it is trivial. And what about the other M's?

S. Masta, I see that if M is infinite, then the system of T_M is not uniquely solvable. The coproduct $T_M + C_{0,1}$ also solves the system of T_M and, since T_M does not contain an internal copy of $C_{0,1}$, T_M is not naturally equivalent to $T_M + C_{0,1}$.

M. That's right. And what about the other M's?

S. Probably, their systems **are** uniquely solvable but it would require some computation.

M. Hence we have

> **Problem 14.**
> Are the systems of functorial equations of the functors $T_M(-) = - \times M$ uniquely solvable for $2 < |M| < \aleph_0$?

S. Masta, in [6], I met a rather interesting functor β. It sends any set X to the set of all ultrafilters on X and, for any map $f\colon X \longrightarrow Y, \beta f$ sends any ultrafilter $\mathcal{U} \in \beta X$ to the ultrafilter $\{Z \subseteq Y \mid f^{-1}(Z) \in \mathcal{U}\}$.

M. This is a quite natural functor, in fact it is the composition

$$\mathbf{Set} \longrightarrow \mathbf{Tych} \longrightarrow \mathbf{Comp} \longrightarrow \mathbf{Set}$$

where the first arrow is the functor which interprets every set as the discrete space, the second arrow is the Čech–Stone compactification functor and the last arrow is the forgetful functor.

S. Do you think, Masta, that the system of this functor is uniquely solvable?

M. It is very nice, my boy, that you formulate interesting questions yourself. Thus we have

> **Problem 15.**
> Is the system of functorial equations of the ultrafilter functor β uniquely solvable?

Next day

S. Masta, I fully solved my Problem 15. The system of functorial equations of the ultrafilter functor β is not uniquely solvable.

M. Really? Tell me, please!

S. The system of β is solved also by the functor $F = (\beta + \beta) \,/\, \mathsf{Id}$ obtained from two copies of β by the identification of their subfunctors of all fixed ultrafilters. These subfunctors are naturally equivalent to Id, so I work with them as with the unique copy of Id in F, i.e., FX is equal to X and two disjoint copies of $\beta X \setminus X$.

M. The functor F really solves the system of functorial equations of β. But why is F not naturally equivalent to β?

S. Let us suppose that $\tau\colon \beta \longrightarrow F$ is a natural equivalence. It must be identical on the copy of Id in β and the copy of Id in F, evidently. Let X be an infinite set, hence τ_X must send $\beta X \setminus X$ bijectively onto two disjoint copies of itself in $FX \setminus X$. Consequently there exist $\mathcal{U}, \mathcal{V}$ in $\beta X \setminus X$, $\mathcal{U} \neq \mathcal{V}$ such that τ_X sends $\mathcal{U}$ on $\mathcal{V}$ in one of the copies of $\beta X \setminus X$. Find disjoint sets $A, B \subseteq X$ such that $A \in \mathcal{U},\ B \in \mathcal{V}$ and let $f\colon X \longrightarrow X$ be a map identical on B and sending all of A to a point $a \in X$. Then $[\beta f](\mathcal{V}) = \mathcal{V}$ while $[\beta f](\mathcal{U}) = a$. And Ff is equal to βf on each of the two copies of $\beta X \setminus X$, hence $[Ff](\mathcal{V}) = \mathcal{V}$. We get

$$Ff \circ \tau_X \neq \tau_X \circ \beta f$$

because $[\tau_X \circ \beta f](\mathcal{U}) = a$, while $[Ff \circ T_X](\mathcal{U}) = [Ff](\mathcal{V}) = \mathcal{V}$.

M. Very good!

S. I had a very fruitful evening yesterday. I also have a general statement about unique solvability.

M. Really? Tell me your statement!

S. If F: **Set** $\longrightarrow$ **Set** has a unique internal copy of C_1 (or $C_{0,1}$) and if F has a proper subfunctor G containing this copy of C_1 (or $C_{0,1}$) as well, then the system of functorial equations of F is not uniquely solvable.

M. How do you prove this ?

S. Let F and G be as above. Let F/G be the functor obtained from F by collapsing G onto the internal copy of C_1 (or $C_{0,1}$). Then the functor H obtained from G and F/G by gluing along their internal copies of C_1 (or $C_{0,1}$) also solves the system of F.

M. It surely solves it. But are you sure that it is not naturally equivalent to the functor F?

S. Oh, Masta, you are right, it could be precisely the same functor as F. I forgot to tell you the assumption about the proper subfunctor G, namely:

> if there is a set X and $a \in FX \setminus GX$ such that every $x \in FX$ outside C_1 (or $C_{0,1}$) is equal to $[Ff](a)$ for suitable $f\colon X \longrightarrow X$ and there exists $y \in GX$ which is still outside C_1 (or $C_{0,1}$),

then the functor H is not naturally equivalent to F.

M. Now, it is correct.

S. And this statement can be applied to the power–set functor P and to its subfunctors $P_{\leq\alpha}$ with $\alpha > 2$.

M. That's right. But it seems to me that a functor $F\colon \mathbf{Set} \longrightarrow \mathbf{Set}$ with $FX = \exp X$, but not naturally equivalent to the power–set functor P, was constructed by V. Koubek a long time ago. I do not know whether he published it.

S. Never mind, Masta. I did it for my own pleasure. There are other interesting functors for which the unique solvability of their systems of functorial equations can be investigated.

> **Problem 16.**
>
> a) Which of the hom-functors $\hom(M, -)$ and their factor-functors $Q_{M,\mathcal{F}}$ have their systems of functorial equations uniquely solvable?
>
> b) Are the systems of functorial equations of B_α with $2 < \alpha$ uniquely solvable?

M. Also operations on functors can be investigated in connection with systems of their functorial equations, e.g., you proved that the system of B_2 is uniquely solvable but the system of $B_2 + B_2$ fails to be uniquely solvable.

S. But the unique solvability of the systems of coproducts of functors does not fail in all cases, e.g., the system of Id is uniquely solvable and the system of $\mathsf{Id} + \mathsf{Id}$ is also uniquely solvable.

M. But what about the converse question?

> **Problem 17.**
> If the system of functorial equations of $F+F$ is uniquely solvable, is the system of F also uniquely solvable?

S. Masta, this must be true! I'll think about it. But, Masta, can you tell me a few words about the remaining references [2], [3], [4], [7], maybe that it could help me.

M. Certainly, my boy. The paper [3] starts from the forms of congruences on the category **Set** and applies them to the behavior of set–functors. Maybe, it could help in the negative reasoning about the systems of functorial equations. The paper [7] investigates preservations of various types of colimits. It contains, e.g., the statement that if a set–functor preserves finite colimits, it preserves also countable colimits. Also ultrafilters closed with respect to infinite intersections and measurable cardinals appear in [7] as well as in [4] where, under the assumption that there is a proper class of measurable cardinals, a non-small functor (= not being a colimit of a set–diagram of hom-functors) preserving all finite limits and finite colimits is constructed. The paper

[2] presents further facts about unattainable cardinals and the size of the corresponding increases and it also investigates contravariant functors **Set** $\longrightarrow$ **Set**. For the problem below, it would be the basic reference.

Problem 18.
Investigate the solvability and the unique solvability of systems of functorial equations also for contravariant functors **Set** $\longrightarrow$ **Set**!

S. Masta, the problems sound promising and I will surely think about them. But I need some time for it.

M. O.K., my boy. In the meantime, I'll finish my reading of "Too Many Cooks"[1], I still don't know who the murderer is.

References

[1] V. Koubek. Set functors, *CMUC*, **12** (1971), 175–195.

[2] V. Koubek. Set functors II, contravariant case, *CMUC*, **14** (1973), 47–59.

[3] V. Koubek and J. Reiterman. Set functors III – monomorphisms, epimorphisms, isomorphisms, *CMUC*, **14** (1973), 441–455.

[4] J. Reiterman. An example concerning set functors, *CMUC*, **12** (1971), 227–233.

[5] V. Trnková. Some properties of set functors, *CMUC*, **10** (1969), 323–352.

[6] V. Trnková, On descriptive classification of set functors I, *CMUC*, **12** (1971), 143–174.

[7] V. Trnková. On descriptive classification of set functors II, *CMUC*, **12** (1971), 345–357.

CMUC = Comment. Math. Univ. Carolinae

Alois Zmrzlina
Matematicky Ustav MFF KU
186 75 Praha 8
Sokolovska 83, Czech Republic
zmrzlina@karlin.mff.cuni.cz

[1] Rex Stout: Too Many Cooks, Pyramid Publication Inc., New York 1963

Contributions and Importance of Professor George E. Strecker's Research

Jürgen Koslowski

ABSTRACT We give an overview of the long and distinguished career of Professor George E. Strecker in the fields of topology and, in particular, categorical topology.

1 Career Overview

Professor George E. Strecker's mathematical career is remarkable in several respects, not the least of which is that he did not start out to become a mathematician. Indeed, he studied electrical engineering and business administration at the University of Colorado, obtaining two B.S. degrees in 1961. But he became another victim of the beauty of mathematics, and so enrolled as a graduate student in mathematics at Tulane University in the fall of that year. Actually, we also have to thank the Soviets for that, since their successful launch of Sputnik in 1957 scared the U.S. Congress into making large sums of money available in the form of National Defense Fellowships. Charles Aull, a teaching assistant in the Department of Applied Mathematics at Colorado (and now at Virginia Tech) encouraged George Strecker to apply for one of those, which he got. The rest, as they say, is history.

George did not just study mathematics at Tulane University. He and his wife Julie (also from Colorado) started a family in New Orleans. Daughter Cheri was born in November of 1961, son Mark in February of 1963.

Mathematically, he soon was seduced by topology. A 1964 colloquium talk by Johannes de Groot from Amsterdam on cotopologies and the Baire category theorem inspired George to work on some aspects of cotopologies. The results he obtained prompted de Groot to suggest that he apply for a Fulbright fellowship and come to Amsterdam. The application was successful, and so George took his young family to Holland for the academic year 1965/66. Most of the research for his dissertation was conducted during this time, including the actual writing of the Ph.D. thesis, entitled "Cotopologies and Generalized Compactness Conditions". Besides that, George also managed to meet M. C. Escher, whose art had fascinated him, as it did many other mathematicians.

When de Groot accepted a part-time distinguished visiting professorship at the University of Florida in Gainesville, teaching one quarter each winter, George came along as a post-doc for the academic year 1966/67. Now the second mathematical seduction took place. By accident, Horst Herrlich from Berlin, Germany, spent the same year in Gainesville as a Visiting Assistant Professor. He was about to become, almost single-handedly, one of the founding fathers of the new discipline of "categorical topology" that pulled together a number of scattered results in topology and put them into a coherent categorical framework (cf., e.g., Figure 3 in [HS97a] concerning the categorical background for the theory of topological reflections and coreflections). His ground-breaking book *Topologische Reflexionen und Coreflexionen* [Her68] appeared as Springer Lecture Notes **78** in 1968, hence must have been in its finishing stage during 1967.

The first meeting of George and Horst in Gainesville proved to be the beginning of a life-long professional partnership and personal friendship. Right away they wrote a joint paper on coreflective sub-categories [HS71a] and submitted it to the *Transactions of the American Mathematical Society.* Inexplicably, there it languished for four years before finally being published in 1971. This was the first of eighteen joint papers over the years, some of which included other collaborators. The last one, so far, appeared in 1997.

Their first joint textbook on category theory [HS73] was conceived during a second stay of Horst in Gainesville in 1969. This book, published in 1973, was highly influential in making category theory accessible for many students. A second edition (this one did the trick for the author) followed in 1979. Another textbook [AHS90], this time a collaborative effort with Jiří Adámek from Prague, took shape in the mid-1980s. What may be perceived as a shift in perspective compared to the first book was in fact a sharpening of a position held by the senior authors from the beginning: category theory, besides its intrinsic beauty and allure, has to serve as a tool for analyzing problems in other mathematical areas. Consequently, the second book emphasized the importance of so-called "concrete categories" of structured sets. These are the ones non-specialists encounter in practice most frequently, and where much conceptual progress had been achieved during the preceding decade, in particular by these three authors. In particular this concerned factorizations systems, not just of single morphisms, but also of sources (i.e., collections of morphisms with common domain), and the interplay of "forgetful functors" with such factorizations. Then many "good" properties of such functors can be specified succinctly and in a uniform fashion. This attitude was reflected in the title "Abstract and Concrete Categories". The subtitle "The Joy of Cats" and the use of the delightful cartoons throughout the book displayed another character trait of the senior authors that everybody who ever had a chance to meet them will immediately have recognized: a playfulness and youth of spirit rarely associated with such a "serious" profession, along with a healthy sense of irreverence, *cf.* [Str01].

George stayed in Gainesville as an Assistant Professor for another four years until 1971. During this time a topology textbook, jointly authored with Philip Nanzetta took shape [NS71], which finally appeared in 1971. When the department head, John Maxfield, left to become chairman and to establish a Ph.D. program at Kansas State University in Manhattan, KS, he offered George a job there. George was not quite ready to leave Gainesville, but eventually took John up on his earlier offer. However, he also was in contact with an attractive group of topologists at the University of Pittsburgh, lead by Jun-Iti Nagata. With close proximity to others interested in topological research, for example Stan Franklin and Oswald Wyler at Carnegie Mellon University, George opted for a one-year stay in Pittsburgh as an Associate Professor to pursue these research opportunities before settling at KSU (a place more suitable for raising children). He has been working there ever since, being promoted to Full Professor in 1977.

August 1971 also saw George's first trip to Prague, for the first meeting on General Topology and its relations to Modern Algebra and Analysis. He and Horst traveled there together. It marked the beginning of very fruitful contacts with the Prague School of mathematicians that led to three extended stays in Prague during sabbaticals and several joint papers, in particular with Jiří Adámek, starting in the second half of the 1970s. The joint work with Jiří Adámek eventually gave rise to the second textbook on abstract and concrete categories [AHS90].

Despite the flourishing of category theory in the 1970s, neither it, nor categorical topology ever was considered a "hot topic" at American universities (or elsewhere), meeting with much irrational resistance. So graduate students were hard to come by. Nevertheless, George directed a number of Ph.D. theses over the years, with a remarkable accumulation during 1971:

- L. W. Smith, "Natural compactifications of lattices", 1969
- T. H. Fay, "Relation theory in categories", 1971
- D. W. Hajek, "Almost Hausdorff extensions", 1971
- J. M. McDill, "Categorical embeddings and linearizations", 1971
- T. L. Goulding, "Regular hereditary subcategories", 1971
- A. C. Melton, "Diafactorization structures", 1980
- T. Titcomb, "Fibers of relatively topological functors", 1980
- J. Koslowski, "Dedekind cuts and Frink ideals in categories", 1986
- G. Castellini, "Closure Operators, epimorphisms and Hausdorff objects", 1986

- W. D. Miller, "Quasi-Heyting algebras: a new class of lattices, and a foundation for non-classical model theory with possible computational applications", 1993
- B. Legan, "On a structure for commutator-finite orthomodular lattices", 1998
- C. Mosbo, (in progress)

2 The Topological Papers

Of course, a strict separation into topological papers and categorical ones is not possible. Nevertheless, some tend to be more topological, and have corresponding classifications in the *Zentralblatt* or *Mathematical Reviews*, once these were introduced.

George's first published paper [vEBMvdS$^+$65] at the Mathematical Center in Amsterdam, a collaboration with P. van Emde Boas, D. Mantel, J. van der Slot and E. Wattel, all students of Johannes de Groot, concerned the k-axiom and about ten of its relatives. Although of purely topological nature, perhaps unwittingly it already pointed towards categorical topology: Since their inception in 1950 by D. Gale [Gal50], k-spaces (also known as compactly generated spaces) had become increasingly important in various people's research, as outlined in the historical overview [HS97a]. In modern terminology, this can be explained by the fact that they form a cartesian closed category, i.e., one with well-behaved function spaces, and may be viewed as the coreflective hull of all compact Hausdorff spaces in ***Haus***, the category of all Hausdorff spaces (and continuous maps). But this was not yet on George's mind in 1966, when in [Str66] he considered topologies in which the non-empty open sets coincide with the dense sets. A fruitful collaboration with E. Wattel ensued, resulting in another five joint papers. The articles [SW66] and [SW67] concerned the possibility of embedding a topological space X into a semi-regular space Y, the latter having a basis of regular open sets. In the first case it was shown that every space X can be embedded into a semi-regular space Y that is the disjoint union of copies of X. In the second paper the possibility of a nowhere dense embedding into a connected semi-regular space that preserved various separation axioms was proved. This implied the possibility of embedding every Hausdorff space into a minimal or *H-closed* one (with no strictly weaker Hausdorff topology).

The principal author of the other three papers with E. Wattel as co-author, [dGSW67], [SWHdG68], and [dGHSW69], was de Groot, although [SWHdG68] also involved Horst, who met de Groot during his first stay in Gainesville. The objective was to study two operators, γ (initially denoted by ϵ) and ρ, on the double power-set of a set X. While γ maps a collection

$\mathcal{C} \subseteq \in^{\mathcal{X}}$ to the collection of all intersections of finite unions of elements of $\mathcal{C}$, and hence is extensive, the collection $\rho\mathcal{C}$ consists of all those subsets $A \subseteq X$, for which the finite intersection property of $\mathcal{C} \cup \{\mathcal{A}\}$ implies that $\mathcal{C} \cup \{\mathcal{A}\}$ has non-empty intersection, i.e., A is *relatively compact* with respect to $\mathcal{C}$. In [dGSW67], the two-element set $\{\gamma, \sigma := \rho^2\}$ is shown to be a monoid with neutral element γ. In [SWHdG68] the collection $\mathcal{D} = \gamma(\mathcal{C} \cup \rho^{\in}\mathcal{C})$ is shown to satisfy $\rho\mathcal{C} = \rho\mathcal{D}$, improving Alexander's result of $\rho\mathcal{C} = \rho\gamma\mathcal{C}$. In particular, for a Hausdorff space $(X, \mathcal{C})$ with $\mathcal{C}$ the collection of closed sets, $\mathcal{D}$ is maximal with this property, and $(X, \mathcal{C})$ is a k-space iff $\mathcal{C} = \mathcal{D}$. In [dGHSW69], higher powers of the compactness operator ρ were used to study "antispaces". These are pairs of topological spaces over the same sets such that the collection of closed sets of the first coincides with the collection of compact subsets of the second space, and vice versa.

The conceptually similar idea of a topology σ (of open sets) on a set X forming a *cotopology* for a given topology τ on X, provided the complements of the closures of some base $\beta \subseteq \tau$ form an open subbase of σ, also is due to de Groot. In [SV69], George and G. Viglino used this notion to characterize minimal Hausdorff spaces: their topology coincides with each of its cotopologies. This leads to a shorter proof of Banaschewski's result concerning the existence of minimal Hausdorff extensions of all semi-regular spaces [Ban61].

The next mainly topological paper [LS72] together with C. T. Liu was originally submitted in December of 1969, and revised in April 1970. It strongly makes use of the categorical insights acquired so far by setting up an analogy between the largest realcompactifications below the Stone–Čech-compactification of a completely regular Hausdorff space X, resulting in the realcompact epi-reflection νX of X, and Frolik's almost realcompactifications [Fro63] below Katětov's H-closed extension of a Hausdorff space Y. However, in the second case continuous maps are not the correct ones to obtain an almost realcompact epi-reflection for every Hausdorff space. They need to be restricted to the continuous *demi-open* ones, which map subsets with non-empty interior to images whose closure has non-empty interior. The idea of changing the maps from the standard ones to those that give the desired elegant result is typical for the way categorical thinking can lead to new insights in an established field.

As an epireflective sub-category of ***Top***, the category ***Haus*** of Hausdorff spaces is closed under limits in ***Top***, but not under colimits. Even filtered colimits (also known as "direct limits") of Hausdorff spaces can fail to be Hausdorff. During his stay in Pittsburgh, George, together with D. W. Hajek, a former colleague from Gainesville, gave two sufficient conditions on filtered diagrams in ***Haus*** to ensure that the colimit in ***Top*** is again a Hausdorff space [HS72a]. These involve an interplay between properties of the maps in the given diagram (being embeddings or relatively open, respectively) and the existence of a cone ("natural source") for the diagram consisting of maps that have a complementary property (being dense or dense embeddings, respectively). The result was presented at the Third Prague Topological Symposium in 1971.

Given George's interest in compactifications and perfect morphisms, as exemplified, e.g., in [Str72], it was no surprise that he was attracted by the compactification, respectively prefection, of continuous surjections between Hausdorff spaces. While such constructions were known, cf. [Why53] and [Cai69], they were unneccessarily complicated. Together with F. A. Delahan, a former colleague from Pittsburgh, George described a simple method to perform these constructions by means of the graph of the function to be improved. While [DS73] from 1973 was just a brief announcement demonstrating the considerable simplifications of this new approach, a more detailed presentation [DS77] was delayed until the Second Symposium on Categorical Topology in Cape Town, 1976.

The last joint paper [HS97b] of George and Horst so far marks a return to purely topological questions, with particular emphasis on the strength of the underlying set theory. It presents a number of topological conditions equivalent to $\mathbf{N}$ being Lindelöf in the absence of the axiom of choice.

3 The Categorical Papers

Horst Herrlich was very interested in charting the progress of the relatively young field of categorical topology over the years [Her83], [HH92], [HH93], [HS97a]. Besides the survey article by Ryosuke Nakagawa [Nak89] and the book by Gerhard Preuss [Pre88], an extensive bibliography has been compiled by Oswald Wyler (available by email from owylernqi.net). George only recently joined Horst in this quest for some historical clarification in a joint effort [HS97a] concerning the period before 1971. It specifically dealt with the developments leading to the notions of topological reflection and coreflection and uncovered some neglected contributions to the development of category theory by general topologists. Particularly noteworthy is perhaps P. Samuel's article [Sam48] from 1948, predating the "official" invention of adjoint functors by D. M. Kan [Kan58] by a full decade. The dependency diagrams of the papers contributing to some of the notions are certainly more complicated than any categorical diagram the author has ever seen (or managed to run through TEX) and portray the difficulty of unraveling the history even of such a limited area of mathematics. Since it is so much easier to forge ahead with new ideas, such efforts of uncovering the far-reaching roots of any particular subject are all the more laudable.

George Strecker's main contributions to the field of categorical topology occurred roughly in the following areas:

- (co)reflective subcategories, various completions and hulls, injectivity;
- factorization structures, not just for morphisms but also for sources, and in particular relative to a functor;

- special types of functors encompassing topological and algebraic features;
- the systematic use of Galois connections induced by simple relations, e.g., for specifying "nice" classes of objects or morphisms, or in the context of categorical closure operators. The success of this method also resulted in attempts to convince others, e.g., computer scientists or general topologists, of its usefulness, less as an ad-hoc solution to any specific problem, but rather as a guiding principle to uncovering previously hidden meaning.

We will describe three distinct periods in George Strecker's research in categorical topology. While there is a bit of overlap, the reader should keep in mind that frequently collaborations with different colleagues were happening roughly at the same time. Moreover, the publication dates of the articles in some cases do not reflect the actual development of ideas.

3.1 The Period 1967–1976

As mentioned before, the first joint article of George and Horst concerning categorical topology [HS71a] resulted from Horst's visit to Gainesville 1966/67, but was delayed in print until 1971. The article concentrates on mono-coreflective subcategories and their characterization in terms of closedness under certain types of limits. It already employs factorization properties of morphisms as an essential tool. In a sequel, [HS72b], these results were applied to study coreflective subcategories of the category ***Top*** of all topological spaces and the category ***Haus*** of all Hausdorff spaces, in both cases with continuous maps as morphisms. Methods of generating such categories as hulls from smaller ones were also investigated.

In 1971 the new discipline of categorical topology also obtained a foothold in the journal market with the launch of *General Topology and its Applications*. The first article of the first issue was written by Horst; it was an overview of the emerging field [Her71]. Another early contribution [HS71b] of Horst and George foreshadowed a line of research that would be fruitful for years to come: the interplay between topology and algebra. Here they combined their by now well-established notion of epi-reflective subcategory with Linton's approach to "varietal" categories in terms of (possibly infinitary) operations and identities [Lin66] to characterize the category $\mathcal{C}$ of compact Hausdorff spaces as a varietal non-trivial epi-reflective subcategory of ***Haus***. Just a year earlier, $\mathcal{C}$ had been shown to be monadic over ***Set*** by Manes [Man69]. The notion of monadicity never quite appealed to both Horst and George, since monadic functors are not closed under composition and may fail to preserve regular epis [Her73] and to detect colimits [Adá77a]. The characterization of these functors by Beck [Bec67] in terms of splitting (or reflexive) coequalizers rather reenforced their distaste for the subject. In their paper [DS73] they introduced the alternative notion of an *algebraic functor* that has a left adjoint and preserves and reflects regular epis.

In terms of the new tools introduced, the article [Str72] that appeared in the *Bulletin of the Australian Mathematical Society* in 1972, although short and lacking detailed proofs, could still be viewed as a landmark. The characterization of epi-reflective subcategories of "nice" categories $\mathcal{C}$ (satisfying suitable completenes and smallness conditions) in terms of stability under formation of products and "sufficiently strong" subobjects was by then well known, one important precursor being the 1966 paper by John Isbell [Isb66]. Correspondingly, epireflective hulls of a given subcategory usually were obtained by first forming products of objects in the subcategory, and then taking appropriate subobjects thereof. Instead, motivated by Horsts's definition of *perfectness* of a class of morphisms with respect to a class of objects [Her72], George in this article began the investigation of the Galois connections induced by two fairly simple relations:

- a morphism f is related to another morphism g, provided that for all morphisms h and k satisfying $h \circ g = f \circ k$ there exists a "diagonal morphism" d (not necessarily unique) with $d \circ g = k$ and $f \circ d = h$;

- a morphism $f\ A \to B$ is related to an object K, provided that for each $g\ A \to K$ there exists some $h\ B \to K$ (not necessarily unique) with $h \circ f = g$, in other words, if K is *injective* with respect to f.

The notion of injectivity would later be of interest in a different context (cf. below). Explicitly requiring uniqueness in both cases led to the well-known notion of *orthogonality.* Notice that Galois connections as an order-theoretic concept are not affected by size problems and hence work for power-conglomerates of all subclasses of certain classes in the same fashion as for power-sets. The Galois connections resulting from the relations above (on the power-conglomerate of $\boldsymbol{Mor}(\mathcal{C})$ in the first case and between this power-conglomerate and that of $\boldsymbol{Ob}(\mathcal{C})$ in the second case) may be combined with the Galois connection induced by the domain-relation and with the intersection operations with the classes of epis or monos in $\mathcal{C}$. This yields very general closure and interior operators. In particular, under the rather weak completeness requirement that every 2-sink (two morphisms with a common codomain) is part of a commutative square, one obtains a Galois connection between the power-conglomerate of $\boldsymbol{Epi}(\mathcal{C})$ and that of $\boldsymbol{Ob}(\mathcal{C})$, such that the corresponding fixed point classes have very nice properties. In particular, if $\mathcal{C}$ is co-wellpowered and has multiple pushouts, forming the closure of a class of objects in this fashion produces its epireflective hull.

The detailed proofs and examples only promised in [Str72] finally appeard in [Str74a], the proceedings of a conference dedicated to the memory of Johannes de Groot, whose untimely death in 1972 had shocked the topology community. While the investigation of epi-reflective subcategories always had depended on the ability to factor arbitrary morphisms into a suitable epi followed by an appropriate mono, the Galois approach now made it possible to treat such factorization systems more abstractly. George started such an investigation in [Str74b], linking it with the specific topological example of the (monotone, light) factorization and deriving some generalizations.

The next conceptual break-through occurred in [Str76], when the first relation between morphisms above was generalized to a relation between morphisms and *sources*, i.e., collections of morphisms with a common domain. [Strangely enough, it was not observed that empty sources correspond to objects, which allows the second relation above between morphisms and objects to be subsumed as well.] Roughly at the same time, Dieter (Nico) Pumplün at the Fernuniversität Hagen and his group seem to have considered similar notions, which led, e.g., to the article [Tho78] by Walter Tholen.

3.2 The Period 1979–1985

It had long been recognized that certain properties of categories were not intrinsic but rather crucially depended on their "forgetful" functors, usually to $\boldsymbol{Set}$, but also to other categories (consider, e.g., topological groups). This led to the notion of "concrete category" over some base-category $\mathcal{X}$, namely a pair $(\mathcal{A}, U)$ where U is a faithful functor from $\mathcal{A}$ to $\mathcal{X}$. Usually U is also required to be *amnestic*, i.e., the induced order on the U-fibres is anti-symmetric, and *uniquely transportable*, i.e., every $\mathcal{X}$-isomorphism $U(A) \to X$ is the U-image of a uniquely determined $\mathcal{A}$-isomorphism.

Just as there had been attempts to characterize concrete categories of an "algebraic"' nature (e.g., in terms of monadic functors U), a number of researchers contributed to the complementary effort of capturing the "topological" nature of a concrete category. Besides his own contributions [Her74c], [Str74], [Her74a] and [Her76], Horst in [Her71] lists the following references as relevant: [Huš64], [Tay65], [Ant66b], [Ant66a], [Wyl71a], [Man72], [Hof72], [Ben73b].

Definition. A functor $T \mathcal{A} \to \mathcal{X}$ is called *topological* or *initially complete*, if every source of the form $\langle f_i\ X \to TA_i \langle_{i \in I}$ has a unique T-initial lift $\langle g_i\ A \to A_i \rangle_{i \in I}$.

Notice that I may be a proper class. It can easily be shown that a topological functor is faithful, amnestic and uniquely transportable. The first of these observations prompted George Strecker and his graduate student Tim Titcomb to consider the possibility of expressing faithfulness itself in terms of the behaviour of suitable sources. This led to the notion of "relatively topological" functor, also suitable to capture other properties of functors, and was presented at the 1981 Symposium on Categorical Algebra and Topology in Cape Town [ST83].

The theory of concrete categories over a fixed base $\mathcal{X}$ is very similar to the theory of partially ordered classes; in fact, the latter arises as a special case of the former when the base is chosen to be the terminal category $\mathbf{1}$ (with precisely one morphism). One case in point is that the notions of initial completeness and of final completeness turn out to be equivalent, just as for a complete lattice it suffices to require either the existence of all infima, or of all suprema.

While the notion of topological functor was driven by a number of topological examples, there were just as many concrete categories, relevant in practice as well, that failed to be topological. Hence just as one likes to embed a partially ordered class meet-densely into a complete one, in general the desire is to find an initially dense embedding (concrete over the base) of a given concrete category into an initially complete one. In addition, the target can be required to be either minimal (MacNeille completion) or universal (with respect to initiality-preserving concrete functors), or maximal.

While for posets such completions always exist, for partially ordered classes size problems may prevent certain completions from existing. The situation is similar in the general case. For small categories, Horst had constructed the corresponding completions in [Her76]. For large categories, the size problems again require a careful analysis. This was carried out by Horst and George together with Jiří Adámek from Prague in [AHS79a]/[AHS79b], where they gave necessary and sufficient conditions for these types of completions to exist in the general case. In particular, for a signature Σ the category of Σ-algebras does not have a largest completion, if Σ contains an operation of arity ≥ 2, or more than one unary operation. The successful collaboration with Jiří Adámek prompted George to spend two months of his 1980 sabbatical in Prague, where he would return again in 1987 and in 1994. As a postscript to these papers, Jiří Adámek and George later completed the picture and showed [AS82] that for signatures consisting of a single unary operation, the largest initial completion of $\boldsymbol{Alg}(\Sigma)$ does always exist over $\boldsymbol{Set}$, but fails to be fibre-small.

The related question arises as to what properties of forgetful functors will guarantee the existence of a MacNeille completion or a universal completion. In particular, this leads to modifications of the concept of topological functor, quite a number of which appeared in the mid-1970s. We only mention two particularly important ones.

Definition. A faithful functor $U\,\mathcal{A} \to \mathcal{X}$ is called *topologically-algebraic*, cf. [Hon74], if every source of the form $\langle f_i\, X \to UA_i \rangle_{i\in I}$ admits a factorization as an $\mathcal{X}$-morphism $e\, X \to UA$ that is U-epi (*i.e.*, $e = Us \circ e$ implies $r = s$), followed by the U-image of an $\mathcal{A}$-source $\langle g_i\, A \to A_i \rangle_{i\in I}$ that is U-initial. Replacing the U-initiality of this source by the semi-initiality of the triple $(e, A, \langle g_i \rangle_{i\in I})$ results in the notion of a *semi-topological functor*, later renamed as *solid*, *cf.* Věra Trnková [Trn75], Rudolf-Eberhard Hoffmann [Hof76a], Walther Tholen [Tho78] and Manfred Wischnewsky [Wis78].

The precise relationship between the strengths of these notions was exhibited in a joint paper by Horst Herrlich, Ryosuke Nakagawa, George Strecker and Tim Titcomb [HNST80]. Although this only appeared in 1980, it is already quoted in [AHS79a]/[AHS79b] and in other papers that appeared in 1979.

A variant of the problem above asks for *reflective* completions, where the concrete embedding is required to have a not neccessarily concrete left adjoint. Concrete categories $(\mathcal{A}, U)$ for which a reflective MacNeille completion exists had been characterized by Hoffmann [Hof78] as well as Tholen and Wischnewsky [TW79] as those for which U is semi-topological (or solid). In [HS79a], Horst and George solved the problem for reflective universal completions: they exist iff U is topologically-algebraic in the sense of Hong [Hon74].

The quest was still on to find a good common generalization of topological functors and of functors of an "algebraic" flavor. The two notions introduced above certainly were reasonable candidates. But in [HS79a] it was also shown that the semi-topological (or solid) functors form the compositive hull of the topologically-algebraic ones, which seemed to tip the scale in favour of the latter concept. However, semi-topological (or solid) functors share a shortcoming of monadic functors insofar as they can fail to preserve regular epis. Tim Titcomb, a graduate student of George's from KSU found a simple counterexample. However, provided all categories involved have (regular epi, monosource) factorizations, those semi-topological (or solid) that do preserve regular epis indeed form the compositive hull of all topological functors and all regular monadic functors (cf. Manes [Man67]), or equivalently, of all topological functors and the algebraic functors of [DS73]. In 1991, Horst and George would return to this problem, together with Till Mossakowski, a graduate student of Horst's [HMS91]. Replacing algebraic functors by essentially algebraic ones, indeed all solid functors constitute the compositive hull.

Already in the mid-1960s, the Prague school of Zdeněk Hedrlin, Ales Pultr and Věra Trnková had devised a particularly nice notion of concrete categories over $\boldsymbol{Set}$, the so-called *functor-structured categories*, *cf.* [HPT67] and [Pul67]. For a functor $F\, \mathcal{X} \to \boldsymbol{Set}$ the category $S(F)$ has pairs (X, A) with $X \in \mathcal{X}$ and $A \subseteq FX$ as objects, while those $\mathcal{X}$-morphisms $f\, X \to Y$ that satisfy $Ff[A] \subseteq B$ serve as $S(F)$-morphisms from (X, A) to (Y, B). Together with Jiří Adámek and Horst, George provided a generalization to arbitrary base categories and, up to isomorphism, identified the corresponding reflective modifications with the fibre-small initially complete categories over the base [AHS79c]. They also characterized the concrete full (reflective, respectively E-reflective) subcategories of functor-structured categories.

The lack of well-behaved function spaces in ***Top*** and ***Haus*** and other categories of interest in analysis had fuelled the investigation of coreflective subcategories in the 1960s. Now this question reappeared in a slightly differnet guise: does a concrete category $(\mathcal{A}, U)$ with finite products admit a full and finally dense concrete embedding into a cartesian closed topological $(\mathcal{B}, V)$ such that the powers of $\mathcal{A}$-objects are initially dense in $(\mathcal{B}, V)$? Some groundwork concerning the existence of such CCT-hulls over the base ***Set*** and with the constraint that constant functions are morphisms had been done by Horst and Louis D. Nel [HN77] as well as by Jiří Adámek and Václav Koubek [AK80]. In [HMS91], Jiří Adámek and George gave the first explicit construction of this CCT-hull. At the 1983 Conference on Categorical Topology in Toledo, OH, George presented a major improvement of the construction [Str84] that did not depend on the base being ***Set***, thus obviating the need for considering constants, but did restrict to the earlier construction, once these constraints were reimposed. Further refinements and a number of concrete examples grew out of a collaboration with Jiří Adámek and another mathematician from Prague, Jan Reitermann [ARS85].

In 1980 Austin Melton finished his Ph.D. under George's supervision with a thesis on diafactorization structures. Together they then took another look at the interplay between factorization structures for sources and epi-reflective subcategories [MS82]. In particular, for any fixed factorization structure $(\mathcal{E}, \mathcal{M})$, the finer ones $(\mathcal{E}', \mathcal{M}')$ with $\mathcal{E}' \subseteq \mathcal{E}$ constitute a complete "lattice" (except for antisymmetry). Those elements that satisfy the cancellation condition that $g \circ f \in \mathcal{E}'$ and $f \in \mathcal{E}$ implies $f \in \mathcal{E}$ bijectively correspond to the full and isomorphisms-closed $\mathcal{E}$-reflective subcategories.

Another part of his 1980 sabbatical George chose to spend at the University of L'Aquila in Italy. Two of the mathematicians there, Eraldo Giuli and Anna Tozzi, had recently become interested in certain aspects of shape theory, in particular the notion of *dense subcategory* put forward by S. Mardešić [Mar78]. Together with George they explored the parallels between E-dense subcategories and E-reflective subcategories [GST83] by introducing *E-dense functors*, a common generalization of dense subcategories and right adjoint functors. Under mild constraints, they also gave characterization theorems for the existence of E-dense hulls.

3.3 The Period 1985–2000

The analogy between ordered sets and concrete categories over a fixed base has been mentioned before. Among those results to be carried over from the order-theoretical setting to the more general case was a gem George and his collaborators previously had overlooked, although it perfectly complemented their previous efforts. This concerned the characterization of locales as the injective objects in the category of meet-semilattices by Günter Bruns and H. Lakser in 1970 [BL70]. And of course, locales are precisely the cartesian closed meet-semilattices. The corresponding result in the setting of concrete categories with concrete products over a fixed base with products was presented by Horst and George at the 1985 Durban (RSA) meeting on classical and categoriecal algebra [HS86]. The concretely cartesian closed topological categories were identified as the injective objects. Replacing the embeddings in the definition of injectivity by other classes of morphisms allowed a number of further topological and algebraic examples to be subsumed under the notion of (generalized) injectivity. This was carried out by Horst and George in a joint paper [BHS86] with Hubertus Bargenda, an assistent to Horst in Bremen. Later George and Jiří Adámek discovered that the approach via injectivity is not limited to the case where the base category has products and the concrete category has them concretely. Their very general result [AS89] was the definitive word on the issue of injectivity in quasicategories of concrete categories over a fixed base.

Besides the issue of the compositive hull of topological and (essentially) algebraic functors [HMS91], another earlier line of research was picked up again and improved upon. Together with Graciela Salicrup from Mexico, Horst and George embarked on a generalization of the topological notions "Hausdorff", "compact", "perfect" and "closed" to more general categorical settings, where the relations between these concepts would stay intact [HSS87]. This required a closer analysis of the interaction between factorization systems for morphisms and (1) strong limit operators, (2) a particular Galois correspondence between classes of objects and classes of morphisms. The latter had been introduced by Dieter (Nico) Pumplün and H. Röhrl in order to deal with issues of convexity [PR85] and was induced by the following "separating" relation between morphisms e and objects Y of a category: any two morphims r and s with codomain Y that satisfy $r \circ e = s \circ e$ agree.

Gabriele Castellini, who was a graduate student of George's along with the author from 1982 until 1986, had studied with Eraldo Giuli from the University of L'Aquila before coming to the United States. In fact, George had met him there during one of his frequent visits and decided to invite him to come to KSU. Together with Dikran Dikranian, Eraldo Giuli had been interested in characterizing epimorphisms in various categories by means of categorical closure operators. Gabriele Castellini's thesis went in a similar direction. The basic set-up concerns categories $\mathcal{X}$ with a factorization structure $(\mathcal{E}, \mathcal{M})$ for sinks; in particular, $\mathcal{M}$ consists of monos and every object has a complete lattice of $\mathcal{M}$-subobjects. A family of closure operators on these lattices that are compatible in the sense that all $\mathcal{X}$-morphisms are continuous (the image of the closure is contained in the closure of the image) constitute a categorical, or "global", closure operator. Actually, it turns out to be useful to relax this notion somewhat by not requiring the individual lattice operations to be idempotent. This allows, e.g., the so-called "ϑ-closure" on a topological space X to be considered.

It adds all those points $x \in X$ to a subset $M \subseteq X$, for which the ordinary closure of some neighborhood of x intersects M. In general, a closure operator F factors a mono $m \in \mathcal{M}$ as $m = m^F \circ m_F$ with both $m_F, m^F \in \mathcal{M}$. The above relaxation means that m^F need not be closed, i.e., $(m^F)_F \cong id$. Similarly, m^F need not be *dense*, i.e., $(m_F)^F \cong id$.

In their first collaboration [CS90], Gabriele Castellini and George combined a construction of Sergio Salbany [Sal75] that related global closure on embeddings in ***Top*** operators with certan subcategories and the Pumplün–Röhrl connection mentioned above for more general concrete categories. Together with the author, they then proceeded to actually factor the Pumplün–Röhrl connection through the lattice of all global closure operators on a given class $\mathcal{M}$ of monos [CKS92a]. This collaboration started during a common visit of Gabriele Castellini and the author to Manhattan, KS, in the summer of 1989 and continued for a couple of years, resulting in a number of joint publications on various aspects of categorical closure operators, always based on a careful analysis of Galois connections arising from suitable relations between either objects or morphisms [CKS92b], [CKS92c], [CKS93], [CKS94a] and [CKS94b]. Actually, among these [CKS93] was written first, in 1990. Its major conceptual advance was the expression of closure operators in terms of the well-known orthogonality relation, generalized to composable pairs. Gabriele Castellini still pursues related issues, while the author has since moved to different areas of category theory.

The presentation of [CKS93] at the University of Wisconsin by George in 1991 raised the interest of the participating topologists in the order-theoretic techniques used. This led to a detailed survey article [EKMS93] on Galois connections and their applications in various areas of mathematics by Austin Melton and George together with the author and Marcel Erné, an order-theorist from Hannover who had supervised the author's 1982 diploma thesis. This collaboration lowered the Erdös-index of Austin Melton, George and the author to 2.

George had been preaching the gospel of Galois connections already to a different audience. When Austin Melton, who had graduated under George's supervision in 1980, accepted a position at the KSU computing sciences department, both of them realized that in many parts of computer science Galois connections were used on an ad hoc basis, without much mathematical background. Having employed this particular tool very effectively in categorical topology, they wanted to make computer scientists aware of this concept they tacitly had been using. This led to a joint paper with Dave Schmidt, also from the KSU computing sciences department, at the 1985 Guildford conference on category theory and computer programming [MSS86] that characterized left adjoint injections and provided some applications of varying degree of difficulty. Austin Melton and George followed this up with a KSU Technical Report [MS86]. Another former Ph.D. student from Gainesville, Jeannie McDill joined Austin Melton and George for a paper that considered Galois connections as the objects of a category and was presented at the 1987 Edinburgh meeting on category theory and computer science [MMS87]. Although the authors could not establish the cartesian closedness of this category, they did succeed, when the morphisms were constrained in a natural way, i.e., for a nice non-full subcategory.

Certain complexity questions in lattice theory motivated the collaboration of George with his colleague Richard J. Greechie from the KSU mathematics department and with Hainal Andréka from Hungary [HJS89]. For every monotone function $f\ P \to Q$ between complete lattices there is a largest residuated (=left adjoint) function $\rho_f\ P \to Q$ below f, known as the *residuated approximation* of f. In case that P is a chain, ρ_f coincides with a certain function $\ell^- f$ known to be computable in polynominal time and useful in cluster analysis [Jan78]. When P is not a chain, $\ell^- f$ may still be defined but can fail to be residuated. The authors introduced another function σ_f, the *shadow* of f, that is again computable in polynomial time and agrees with ρ_f in case that Q is completely distributive. In fact, completely distributive lattices Q are characterized by the fact that every monotone function with codomain Q has a residuated shadow.

As useful as Galois connections turned out to be in computer science, in certain situations they did not quite fit. Sometimes, partially ordered sets P and Q would occur together with monotone functions f and g from P to Q and back such that both, $f \circ g$ and $g \circ f$ were closure operations. In case of a Galois connection one would have one closure operation and one interior operation. This prompted Austin Melton and George together with a particularly bright graduate Student at Kansas State University at the time, Bernd S. W. Schröder, to investigate this new type of situation, which they playfully christenend “Lagois connections” [MSS94]. This paper also took an unusually long time to appear in print. Before it did, the same three authors discovered a common generalization of both Galois connections and Lagois connections in a 1982 paper by Henry Crapo [Cra82]. This opened up a whole hierarchy of concepts, which they analyzed in [MSS92].

In a paper published in 1994, Horst and Walter Meyer [HM94] rediscovered and slightly generalized a notion that Ross Street had discussed a decade earlier in a somewhat different context [Str84]: a generalization of sinks (i.e., collections of morphisms with common codomain) to families of morphisms $f_{i,j}\ A_i \to B_j$, where i ranges over a possibly proper class I, while j ranges over a set J (Street required J to be finite). These may be viewed either as a set-indexed source of sinks with common domain $\langle A_i \rangle_{i \in I}$, or as a sink of small sources with common codomain $\langle B_j \rangle_{j \in J}$. Horst and Walter Meyer called these families *flows* and used them as a tool for investigating the connections between completeness properties of a category and the existence of certain factorization structures. For the 1998 Antwerp meeting in honor of Horst Herrlich's 60th birthday (about 6 months earlier), George applied the well-honed technique of relativization with respect to a functor to this concept [Str00a], leading to nice characterizations of right adjoint functors with domains satisfying various completeness conditions. Mindful of Dieter (Nico) Pumplün's mild criticism of [HM94] in his Zentralblatt review Zbl. 810.18002 for not considering the dual notion of flows, George was eager to remedy the situation and duly introduced the required concept: wolfs!

George has since also contributed a chapter on (E, M)-factorization systems in categories to the recent Encyclopedia of Mathematics [Str00b]

4 Some Personal Remarks

I had the good fortune to meet both George and Horst in 1981 at a Summer School organized by the Studienstiftung des deutschen Volkes in Völz am Schlern (northern Italy). After a seminar on the Stone-duality for Boolean algebras in 1979/80 at the University of Hannover, which was not categorical but used a bit of the terminology, I had sought out the second edition of their first book [HS73] and was immediately hooked on this beautiful subject. So the opportunity of meeting both authors in person was not to be missed, even if it meant traveling to Völz directly after the two-week *Symposium on Ordered Sets* in Banff, Alberta, which I had attended with Marcel Erné from the University of Hannover. So I flew from Calgary via Toronto (brief excursion to the CN Tower between flights) to Frankfurt, took a train to Munich, from where the students were taken to Völz by bus. The Summer School course organized by George and Horst dealt with paradoxes and antinomies, and one of the highlights, typical of the light-hearted spirit both organizers managed to spread around, was the limerick contest. I am grateful for the opportunity to dedicate my contribution to that contest to both George and Horst:

A hangman with bleen-grueish eyes
And a non-black non-raven that flies
Though of quite different kind
Both agree that they find
Gödel numbers especially nice.

This Summer school led to an invitation by George for me to come as a teaching assistant to Kansas State University (KSU) in 1982, after finishing my M.A. (Diploma) in Hannover. What initially was intended as a one-year stay turned into a full-fledged Ph.D. program, successfully terminated in 1986. During those years, and during a later two-year stay at KSU as Visiting Assistant Professor (1990–92, first in the department of mathematics, then in the department of computing and information sciences), I was able to observe and appreciate all three, George the mathematician, George the teacher and George the person. One particularly useful lesson to learn was the value of serendipity, in particular in the face of adversity. From the beginning I was treated almost as a member of the Strecker family, and I have very fond memories of my time in Manhattan, KS, despite its reputation that, while not being the end of the world, one can see it from there.

Acknowledgement

George Strecker's and Horst Herrlich's help in sorting the chronology of some of the events made this article possible. Also, Sonja Melton's help in proofreading was invaluable. Of course, any remaining mistakes or misrepresentations are the responsibility of the author.

In the following lists of references, George Strecker's articles have been separated according to their mainly topological or categorical content. They also have been sorted roughly chronologically by the date of publication.

5 George E. Strecker's Topological Articles

[vEBMvdS+65] P. van Emde Boas, D. Mantel, J. van der Slot, George E. Strecker and E. Wattel. De onderlinge afhankelijkheid van een aantal topologische axioma's die verband houden met het k-axioma (The underlying connections between topological axioms similar to the k-axiom). Technical Report ZW1965-018, Math. Centrum Amsterdam Afd. Zuivere Wisk., 1965.

[SW66] George E. Strecker and E. Wattel. A coherent embedding of an arbitrary topological space in a semi-regular space. Technical Report ZW1966-006, Math. Centrum Amsterdam Afd. Zuivere Wisk., 1966. (Zbl. 148.43003, MR 39#923)

[Str66] George E. Strecker. On topologies congruent with their classes of dense sets. Technical Report ZW1966-019, Math. Centrum Amsterdam Afd. Zuivere Wisk., 1966.

[SW67] George E. Strecker and E. Wattel. On semi-regular and minimal Hausdorff embeddings. *Nederl. Akad. Wetensch. Indag. Math.*, (29):234–237, 1967. (Zbl. 148.43004/5, MR 35#2261)

[dGSW67] Johannes de Groot, George E. Strecker, and E. Wattel. The compactness operator in general topology. In *General Topology and its Relations to Modern Analysis and Algebra, II (Proc. Second Prague Topological Sympos., 1966)*, pages 161–163. Academia, Prague, 1967. (Zbl. 165.25301, MR 38#657)

[SWHdG68] George E. Strecker, E. Wattel, Horst Herrlich, and Johannes de Groot. Strengthening Alexander's subbase theorem. *Duke Math. J.*, 35:671–676, 1968. (Zbl. 169.25001, MR 37#6912)

[HS68] Horst Herrlich and George E. Strecker. H-closed spaces and reflective subcategories. *Math. Ann.*, 177:302–309, 1968. (Zbl. 157.29104, MR 38#2744)

[SV69] George E. Strecker and G. Viglino. Cotopology and minimal Hausdorff spaces. *Proc. Amer. Math. Soc.*, 21:569–574, 1969. (Zbl. 175.49404, MR 39#2115)

[dGHSW69] Johannes de Groot, Horst Herrlich, George E. Strecker, and E. Wattel. Compactness as an operator. *Compositio Math.*, 21:349–375, 1969. (Zbl. 186.55902, MR 41#4490)

[LS72] C. T. Liu and George E. Strecker. Concerning almost realcompactifications. *Czechoslovak Math. J.*, 22:181–190, 1972. (Zbl. 247.54024, MR 46#2644)

[HS72a] D. W. Hajek and George E. Strecker. Direct limits of Hausdorff spaces. In *General topology and its relations to modern analysis and algebra, III (Proc. Third Prague Topological Sympos., 1971)*, pages 165–169, Prague, 1972. Academia. (Zbl. 306.54017, MR 50#8410)

[DS73] F. A. Delahan and George E. Strecker. A simplified approach to the compactification of mappings. *Bull. Amer. Math. Soc.*, 79:1030–1032, 1973. (Zbl. 268.54014, MR 47#7685)

[DS77] F. A. Delahan and George E. Strecker. Graphic extensions of mappings. *Quaestiones Math.*, 2:401–417, 1977. (Zbl. 365.54004, MR 58#2692)

[HS97b] Horst Herrlich and George E. Strecker. When is **N** Lindelöf? *Comment. Math. Univ. Carolinae*, 38(3):553–556, 1997. (Zbl. 938.54008, MR 99f:99c:03070)

6 George E. Strecker's Categorical Articles

[HS71a] Horst Herrlich and George E. Strecker. Coreflective subcategories. *Trans. Amer. Math. Soc.*, 157:205–226, 1971. (Zbl. 224.18003, MR 43#6281)

[HS71b] Horst Herrlich and George E. Strecker. Algebra $\bigcap$ topology = compactness. *General Topology and Appl.*, 1:283–287, 1971. (Zbl. 231.18007, MR 46#7349)

[HS72b] Horst Herrlich and George E. Strecker. Coreflective subcategories in general topology. *Fund. Math.*, 73:199–218, 1972. (Zbl. 231.18008, MR 46#1872)

[Str72] George E. Strecker. Epireflection operators vs perfect morphisms and closed classes of epimorphisms. *Bull. Austral. Math. Soc.*, 7:359–366, 1972. (Zbl. 2242.18004, MR 48#368)

[Str74a] George E. Strecker. On characterizations of perfect morphisms and epireflective hulls. In *TOPO 72—general topology and its applications (Proc. Second Pittsburgh Internat. Conf., Pittsburgh, Pa., 1972; dedicated to the memory of Johannes H. de Groot)*, pages 468–500. Lecture Notes in Math., Vol. 378, Berlin – Heidelberg – New York, 1974. Springer-Verlag. (Zbl. 289.18004, MR 51#1709)

[Str74b] George E. Strecker. Component properties and factorizations. In *Topological structures (Proc. Sympos. in honour of Johannes de Groot (1914–1972), Amsterdam, 1973)*, number 52 in Math. Centre Tracts, pages 123–140, Amsterdam, 1974. Math. Centrum. (Zbl. 293.54010, MR 50#8421)

[Str76] George E. Strecker. Perfect sources. In *Categorical topology (Proc. Conf., Mannheim, 1975)*, pages 605–624. Lecture Notes in Math., Vol. 540, Berlin – Heidelberg – New York, 1976. Springer-Verlag. (Zbl. 338.54007, MR 56#9479)

[HS79a] Horst Herrlich and George E. Strecker. Semi-universal maps and universal initial completions. *Pacific J. Math.*, 82(2):407–428, 1979. (Zbl. 379.18008/414.18005, MR 81f:18005)

[AHS79a] Jiří Adamek, Horst Herrlich, and George E. Strecker. Least and largest initial completions. I. *Comment. Math. Univ. Carolinae*, 20:43–58, 1979. (Zbl. 404.18005, MR 80f:18009)

[AHS79b] Jiří Adamek, Horst Herrlich, and George E. Strecker. Least and largest initial completions. II. *Comment. Math. Univ. Carolinae*, 20:59–77, 1979. (Zbl. 404.18006, MR 80f:18009)

[AHS79c] Jiří Adámek, Horst Herrlich, and George E. Strecker. The structure of initial completions. *Cahiers Topologie Géom. Différentielle*, 20(4):333–352, 1979. (Zbl. 436.18001, MR 81d:18009)

[HS79b] Horst Herrlich and George E. Strecker. Algebra $\bigcup$ topology. In *Categorical topology (Proc. int. Conf., Berlin, 1978)*, volume 719 of *Lecture Notes in Math.*, pages 150–156, Berlin – Heidelberg – New York, 1979. Springer-Verlag. (Zbl. 399.18003, MR 80j:18013)

[HS79c] H. Herrlich and George E. Strecker. Semi-universal maps and universal initial completions. In *Structure of topological categories (Proc. Conf., Univ. Bremen, Bremen, 1978)*, pages 75–108, Bremen, 1979. Univ. Bremen. (MR 82b:18004)

[HNST80] Horst Herrlich, R. Nakagawa, George E. Strecker, and Tim Titcomb. Equivalence of topologically-algebraic and semi-topological functors. *Canad. J. Math.*, 32(1):34–39, 1980. (Zbl. 435.18002, MR 81f:18008)

[AS81] Jiří Adamek and George E. Strecker. Construction of cartesian closed topological hulls. *Comment. Math. Univ. Carolinae*, 22(2):235–254, 1981. (Zbl. 457.18006, MR 83a:18017)

[AS82] Jiří Adamek and George E. Strecker. On the largest initial completion of categories of algebras. In *Categorical aspects of topology and analysis (Ottawa, Ont., 1980)*, volume 915 of *Lecture Notes in Math.*, pages 1–15, Berlin – Heidelberg – New York, 1982. Springer-Verlag. (Zbl. 479.18001, MR 83g:18002)

[MS82] A. Melton and George E. Strecker. On the structure of factorization structures. In *Category theory (Gummersbach, 1981)*, volume 962 of *Lecture Notes in Math.*, pages 197–208, Berlin – Heidelberg – New York, 1982. Springer-Verlag. (Zbl. 502.18001, MR 84b:18001)

[GST83] Eraldo Giuli, George E. Strecker, and Anna Tozzi. On e-dense hulls and shape theory. In *General topology and its relations to modern analysis and algebra, V (Prague, 1981)*, volume 3 of *Sigma Ser. Pure Math.*, pages 233–238, Berlin, 1983. Heldermann Verlag. (Zbl. 509.18006, MR 84g:18003)

[ST83] George E. Strecker and Tim Titcomb. Faithfulness vs. topologicity. *Quaestiones Math.*, 6:255–264, 1983. (Zbl. 525.18001, MR 84f:18001)

[Str84] George E. Strecker. On cartesian closed topological hulls. In *Categorical topology (Toledo, Ohio, 1983)*, volume 5 of *Sigma Ser. Pure Math.*, pages 523–539, Berlin, 1984. Heldermann Verlag. (Zbl. 548.18006, MR 87f:18006)

[ARS85] Jiří Adamek, J. Reiterman, and George E. Strecker. Realization of cartesian closed topological hulls. *Manuscripta Math.*, 53(1-2):1–33, 1985. (Zbl. 573.54006, MR 86m:18005)

[MSS86] A. Melton, D.A. Schmidt, and George E. Strecker. Galois connections and computer science applications. In *Category theory and computer programming (Guildford, 1985)*, volume 240 of *Lect. Notes in Comput. Sci*, pages 299–312, Berlin – Heidelberg – New York, 1986. Springer-Verlag. (Zbl. 622.06004, MR 88h:68059)

[HS86] Horst Herrlich and George E. Strecker. Cartesian closed topological hulls as injective hulls. *Quaestiones Math.*, 9(1-4):263–280, 1986. (Zbl. 614.18003, MR 88a:18008)

[MS86] A. Melton and George E. Strecker. Structures supporting Galois connections. Technical Report CS-86-1, Kansas State University, 1986.

[BHS86] H. Bargenda, Horst Herrlich, and George E. Strecker. Concrete categories and injectivity. In *Mathematical foundations of programming semantics (Manhattan, KS, 1985)*, volume 239 of *Lect. Notes in Comput. Sci*, pages 42–52, Berlin – Heidelberg – New York, 1986. Springer-Verlag. (Zbl. 597.18008, MR 88a:18009)

[HSS87] Horst Herrlich, G. Salicrup, and George E. Strecker. Factorizations, denseness, separation, and relatively compact objects. *Topology Appl.*, 27(2):157–169, 1987. (Zbl. 629.18003, MR 89e:18001)

[MMS87] J.M. McDill, A.C. Melton, and George E. Strecker. A category of Galois connections. In *Category theory and computer science (Edinburgh, 1987)*, volume 283 of *Lect. Notes in Comput. Sci*, pages 290–300, Berlin – Heidelberg – New York, 1987. Springer-Verlag. (Zbl. 649.18003, MR 89e:18009)

[AS89] Jiří Adámek and George E. Strecker. Injectivity of topological categories. *Algebra Universalis*, 26(3):284–306, 1989. (Zbl. 685.18007, MR 91e:18005)

[HJS89] Andréka H., Greechie R. J., and George E. Strecker. On residuated approximations. In *Categorical Methods in Computer Science with Aspects from Topology*, volume 393 of *Lect. Notes in Comput. Sci*, pages 333–339, Berlin – Heidelberg – New York, 1989. Springer-Verlag. (MR 91e:06001)

[CS90] Gabriele Castellini and George E. Strecker. Global closure operators vs. subcategories. *Quaestiones Math.*, 13(3-4):417–424, 1990. (Zbl. 733.18001, MR 92h:18001)

[HMS91] Horst Herrlich, T. Mossakowski, and George E. Strecker. Algebra ∪ topology. In *Category theory at work (Bremen, 1990)*, volume 18 of *Res. and Expo. in Math.*, pages 137–148, Berlin, 1991. Heldermann Verlag. (Zbl. 748.18002, MR 92m:18001)

[CKS92a] Gabriele Castellini, Jürgen Koslowski, and George E. Strecker. A factorization of the Pumpluen–Roehrl connection. *Topology Appl.*, 44(1-3):69–76, 1992. (Zbl. 774.18002, MR 93f:18007)

[CKS92b] Gabriele Castellini, Jürgen Koslowski, and George E. Strecker. Categorical closure operators via Galois connections. In W. Gähler et. al., editor, *Recent developments of General Topology and its Applications*, Mathematical Research, pages 72–79, Berlin, 1992. Akademie-Verlag. (Zbl. 793.18001, MR 94e:06004)

[CKS92c] Gabriele Castellini, Jürgen Koslowski, and George E. Strecker. Hereditary and modal closure operators. In R. A. G. Seely, editor, *Category Theory 1991, Proceedings of the Summer International Meeting*, number 13 in C.M.S. Conference Proceedings, pages 111–132, Providence, RI, 1992. Amer. Math. Soc. (Zbl. 796.18002, MR 93k:18005)

[MSS92] A. Melton, Bernd S. W. Schröder, and George E. Strecker. connections. In *MFPS 1991, Proceedings of the International Conference 1991*, number 598 in Lect. Notes in Comput. Sci, pages 493–506, Berlin – Heidelberg – New York, 1992. Springer-Verlag. (MR 94g:06007)

[CKS93] Gabriele Castellini, Jürgen Koslowski, and George E. Strecker. Closure operators and polarities. In *Papers on general topology and applications (Madison, WI, 1991)*, pages 38–52, New York, 1993. New York Acad. Sci. (Zbl. 815.18001, MR 95a:18001)

[EKMS93] M. Erné, Jürgen Koslowski, A. Melton, and George E. Strecker. A primer on Galois connections. In *Papers on general topology and applications (Madison, WI, 1991)* (Ed. Aaron R. Todd), volume 704 of *Annals of the New York Academy of Sciences*, pages 103–125, New York, 1993. New York Acad. Sci. (Zbl. 809.06006, MR 95g:06011)

[CKS94a] Gabriele Castellini, Jürgen Koslowski, and George E. Strecker. Regular closure operators. *Appl. Categ. Structures*, 2(3):219–244, 1994. (Zbl. 801.18003, MR 95h:18001)

[CKS94b] Gabriele Castellini, Jürgen Koslowski, and George E. Strecker. An approach to a dual of regular closure operators. *Cahiers Topologie Géom. Différentielle Catégoriques*, 35(2):109–128, 1994. (Zbl. 801.18002, MR 95f:18003)

[MSS94] Austin Melton, Bernd S.W. Schroeder, and George E. Strecker. Lagois connections – a counterpart to Galois connections. *Theoret. Comput. Sci.*, 136(1):79–107, 1994. (Zbl. 873.68053, MR 96f:06006)

[HS97a] Horst Herrlich and George E. Strecker. Categorical topology – its origins, as exemplified by the unfolding of the theory of topological reflections and coreflections before 1971. In *Handbook of the history of general topology, Vol. 1*, pages 255–341, Dordrecht, 1997. Kluwer Acad. Publ. (Zbl. 906.54001, MR 99f:54003)

[Str00a] George E. Strecker. Flows with respect to a functor. *Appl. Categ. Structures*, 8(3):559–577, 2000.

[Str00b] George E. Strecker. *E-M-factorization systems in categories*, Encyclopedia of Mathematics, pages 374–375. Kluwer, Amsterdam, 2000.

[Str01] George E. Strecker. 10 rules for surviving as a mathematician and teacher This volume.

7 George E. Strecker's Monographs

[Str80a] George E. Strecker. *Introduction to* (E, M)*-Categories and* (E, M)*-Functors.* Consiglio Nazionale dello Ricerche, Rome, 1980.

[Str80b] George E. Strecker. *An Introduction to Categorical Methods.* Wiskundig Seminarium, Amsterdam, 1980.

[Str81a] George E. Strecker. *Keeping Dimensions Straight*, volume 564 of *UMAP Monograph.* EDC Inc., Boston, 1981.

[Str81b] George E. Strecker. *Initial Completions of Concrete Categories.* Consiglio Nazionale dello Ricerche, Rome, 1981.

[Str82] George E. Strecker. *Round Numbers*, volume 578 of *UMAP Monograph.* EDC Inc., Boston, 1982. UMAP Journal III No. 4 (1982),425–454.

8 George E. Strecker's Books

[NS71] Philip Nanzetta and George E. Strecker. *Set theory and topology.* Bogden & Quigley, Tarrytown-on-Hudson, N.Y., 1971. (Zbl. 214.00207, MR 57#120)

[HS73] Horst Herrlich and George E. Strecker. *Category theory. An introduction.* Allyn and Bacon Inc., Boston, 1973. 2nd edition: Heldermann Verlag, Berlin, 1979. (Zbl. 265.18001/437.18001, MR 57#120/81e:18001)

[AHS90] Jiřı Adámek, Horst Herrlich, and George E. Strecker. *Abstract and concrete categories. The joy of cats.* John Wiley & Sons Inc., New York, 1990. (Zbl. 695.18001, MR 91h:18001)

9 Other References

[Adá77a] Jiří Adámek. Colimits of algebras revisited. *Bull. Austral. Math. Soc.*, 17:433–450, 1977.

[AK80] Jiří Adámek and V. Koubek. Cartesian closed initial completions. *General Topology Appl.*, 11:1–16, 1980.

[Ant66a] P. Antoine. Etude elementaire des catégories d'ensembles structures. *Bull. Soc. Math. Belg.*, 18:142–164, 1966.

[Ant66b] P. Antoine. Extension minimale de la catégorie des espaces topologiques. *C. R. Acad. Sci., Paris, Ser. A*, 262:1389–1392, 1966.

[Ban61] Bernhard Banaschewski. Über Hausdorffsch-minimale Erweiterung von Räumen. *Arch. Math. (Basel)*, 12:355–365, 1961.

[Bec67] Jon M. Beck. *Triples, algebras and cohomology.* PhD thesis, Columbia University, 1967.

[Ben73b] H. L. Bentely. T-categories and some representation theorems. *Portugal. Math.*, 32:201–222, 1973.

[BL70] G. Bruns and H. Lasker. Injective hulls of semilattices. *Canad. Bull. Math.*, 13:115–118, 1970.

[Cai69] G. L. Jr. Cain. Compactification of mappings. *Proc. Amer. Math. Soc.*, 23:298–303, 1969.

[Cra82] Henry Crapo. Ordered sets: retracts and connections. *J. Pure Appl. Algebra*, 23:13–28, 1982.

[Fro63] Z. Frolik. A Generalization of Realcompact Spaces. *Czechoslovak Math. J.*, 13:127–137, 1963.

[Gal50] D. Gale. Compact sets of functions and function rings. *Proc. Amer. Math. Soc.*, 1:303–308, 1950.

[Her68] Horst Herrlich. *Topologische Reflexionen und Coreflexionen.* Number 78 in LNM. Springer-Verlag, Berlin–Heidelberg–New York, 1968.

[Her71] Horst Herrlich. Categorical topology. *General Topology and Appl.*, 1:1–15, 1971.

[Her72] Horst Herrlich. A generalization of perfect maps. In *General topology and its relations to modern analysis and algebra, III (Proc. Third Prague Topological Sympos., 1971)*, pages 187–191, Prague, 1972. Academia.

[Her73] Horst Herrlich. Topological structures. In P. C. Baayen, editor, *Topological Structures*, volume 52 of *Mathematical Centre Tracts*, pages 59–122, Amsterdam, 1974, November 1973. Wiskundig Genootschap of the Netherlands, Mathematisch Centrum.

[Her74a] Horst Herrlich. Cartesian closed topological categories. *Math. Colloq. Univ. Cape Town*, 9:1–16, 1974.

[Her74c] Horst Herrlich. Topological functors. *General Topology and Appl.*, 4:125–142, 1974.

[Her76] Horst Herrlich. Initial completions. *Math. Z.*, 150:101–110, 1976.

[Her83] Horst Herrlich. Categorical topology 1971–1981. In J. Novák, editor, *General topology and its relations to modern analysis and algebra, V (Proc. Fifth Prague Topological Sympos., 1981)*, volume 3 of *Sigma Series in Pure Mathematics*, pages 279–383, Berlin, 1983. Heldermann Verlag.

[HH92] Horst Herrlich and M. Hušek. Categorical Topology. In M. Hušek and J. van Mill, editors, *Recent Progress in General Topology*, pages 369–403, Amsterdam, 1992. Elsevier.

[HH93] Horst Herrlich and M. Hušek. Some open categorical problems in t*op*. *Appl. Categ. Structures*, 1:1–14, 1993.

[HM94] Horst Herrlich and W. Meyer. Factorization of flows and completenes of categories. *Quaestiones Math.*, 17:1–11, 1994.

[HN77] Horst Herrlich and Louis D. Nel. Cartesian closed topological hulls. *Proc. Amer. Math. Soc.*, 62:215–222, 1977.

[Hof72] Rudolph-Eberhard Hoffmann. *Die kategorielle Auffassung der Initial- und Finaltopology.* PhD thesis, Universität Bochum, 1972.

[Hof76a] Rudolf-Eberhard Hoffmann. Semi-identifying lifts and a generalization of the duality theorem for topological functors. *Math. Nachr.*, 74:295–307, 1976.

[Hof78] Rudolf-Eberhard Hoffmann. Note on semi-topological functors. *Math. Z.*, 160:69–74, 1978.

[Hon74] Y. H. Hong. *Studies on categories of universal topological algebras.* PhD thesis, McMaster University, 1974.

[HPT67] Z. Hedrlin, A. Pultr, and V. Trnková. Concerning a categorical approach to topological and algebraic categories. In *General topology and its relations to modern analysis and algebra, II (Proc. Second Prague Topological Sympos., 1966)*, pages 176–181, Prague, 1967. Academia.

[Huš64] M. Hušek. S-categories. *Comment. Math. Univ. Carolinae*, 5:37–46, 1964.

[Isb66] John R. Isbell. Structure of categories. *Bull. Amer. Math. Soc.*, 72:619–655, 1966.

[Jan78] M. F. Janowitz. Working report: Seminar on distribution function clustering. preprint, 1978.

[Kan58] D. M. Kan. Adjoint functors. *Trans. Amer. Math. Soc.*, 87:294–329, 1958.

[Lin66] F. E. J. Linton. Some Aspects of Equational Categories. In S. Eilenberg et al., editor, *Proceedings of the Conference on Categorical Algebra (La Jolla, CA, 1965)*, pages 84–94, New York, 1966. Springer-Verlag.

[Man67] E. G. Manes. *A triple miscellany.* PhD thesis, Wesleyan University, 1967.

[Man69] E. G. Manes. A triple theoretic construction of compact algebras. In B. Eckmann, editor, *Seminar on Triples and Categorical Homology Theory*, Lecture Notes in Mathematics, pages 91–118. Springer-Verlag, Berlin – Heidelberg – New York, 1969.

[Man72] E. G. Manes. A pullback theorem for triples in a lattice fibering with applications to algebra and analysis. *Algebra Universalis*, 2:7–17, 1972.

[Mar78] S. Mardešić. Lecture notes on the foundations of shape theory. Technical report, University of Kentucky, 1978.

[Nak89] R. Nakagawa. Categorical Topology. In K. Morita and J. Nagata, editors, *Topics in General Topology*, pages 563–623, Amsterdam, 1989. Elsevier.

[PR85] D. Pumplün and H. Röhrl. Separated totally convex spaces. *Manuscripta Math.*, 50:145–183, 1985.

[Pre88] Gerhard Preuss. *Theory of Topological Structures.* Mathematics and its Applications. D. Reidel, Dordrecht, 1988.

[Pul67] A. Pultr. On selecting of morphisms among all mappings between underlying sets of concrete categories and realizations of these. *Comment. Math. Univ. Carolin.*, 8:53–83, 1967.

[Sal75] S. Salbany. Reflective subcategories and closure operators. In E. Binz and H. Herrlich, editors, *Categorical Topology*, volume 540 of *LNM*, pages 548–565, Mannheim, 1975. Springer-Verlag, Berlin – New York, 1976.

[Sam48] P. Samuel. On universal mappings and free topological groups. *Bull. Amer. Math. Soc.*, 54:591–598, 1948.

[Str74] George E. Strecker. Topological structures. In *Topological structures (Proc. Sympos. in honour of Johannes de Groot (1914–1972), Amsterdam, 1973)*, number 52 in Math. Centre Tracts, pages 59–122, Amsterdam, 1974. Math. Centrum.

[Str84] Ross Street. The family approach to total cocompleteness and toposes. *Trans. Amer. Math. Soc.*, 284:355–369, 1984.

[Tay65] J. C. Taylor. Weak families of maps. *Canad. Math. Bull.*, 8:771–781, 1965.

[Tho78] Walter Tholen. Factorizations of cones along a functor. *Quaestiones Math.*, 2(1–3):335–353, 1977/78.

[TW79] Walter Tholen and Manfred B. Wischnewsky. Semi-topological functors II, External characterization. *J. Pure Appl. Algebra*, 15:75–92, 1979.

[Trn75] Vera Trnková. Automata and categories. In *Math. Found. Comput. Sci. (4th Symp., Marianske Lazne 1975)*, volume 32 of *Lecture Notes in Computer Science*, pages 138–152, Berlin – Heidelberg – New York, 1975. Springer-Verlag.

[Why53] G. T. Whyburn. Compactness of certain mappings. *Trans. Amer. Math. Soc.*, 74:344–350, 1953.

[Wis78] Manfred B. Wischnewsky. A lifting theorem for right adjoints. *Cahiers Topologie Géom. Différentielle*, 19:155–168, 1978.

[Wyl71a] Oswald Wyler. On the categories of general topology and topological algebra. *Arch. Math. (Basel)*, 22:7–17, 1971.

Jürgen Koslowski
Institut für Theoretische Informatik
Technische Universität Braunschweig
Braunschweig, Germany
`koslowj@iti.cs.tu-bs.de`

10 Rules for Surviving as a Mathematician and Teacher

George E. Strecker

I am indeed honored that a gathering was organized to celebrate my sixtieth birthday. When I turned forty, Horst Herrlich told me that at that point I was entitled to wax philosophical[1]. So now I feel that I am more than entitled — I'm actually obliged to do so. So here goes ...

As to format, I would like to take a page out of the book of Gian-Carlo Rota who recently wrote (on a similar occasion) "Ten Lessons I Wish I Had Been Taught" ([1], [2], [3], [4]) and impart to you — especially those of you who are closer to the beginning of a career as a mathematician and teacher — some fatherly (or even grandfatherly) advice in the form of rules. Those of you who are farther along in your careers might read them just to see if you agree.

Rule 1. *It probably doesn't matter.*
Namely, whatever you think is essential or important probably isn't. You don't need to stew endlessly over whether to publish in Journal X rather than Journal Y — or whether to prove your main theorem with a string of lemmas that makes the principal proof a slick one-liner or to grind out all the details in one long impressively complex proof — or whether to come up with only seven pieces of sage advice, rather than (the more standard) ten, — or whether to wear that tie with birds or the one with fish to a grand occasion. In almost every instance it just doesn't matter, and you would be better off spending your time proving more results — or even just improving your tennis serve or repairing your bicycle or listening to music or balancing your check book ...

Rule 2. *No one is trying to steal your results or ideas.*
When others seem interested as you talk to them, they are trying to see how what you tell them will help them solve their particular pressing problems. Remember, that is what you are trying to do when you are particularly interested in what others say or write. They are no different. So there is no need to keep your research notes under lock

[1] I believe that that insight was due to his having turned forty half a year earlier.

and key or not to be forthcoming in conversations. It is rare that the good feelings you get from being called "the person who first proved the truth or falsity of the Smurdley Conjecture" are better than the "high" that you get when proving or disproving it in the first place. Indeed telling others what you know usually leads them to tell you what they know and this is typically where joint publications (and often satisfying friendships) come from. One reason that some are paranoid concerning this issue is that many of us are careless in not giving proper credit for the genesis of our ideas and results. Always be generous with credit and with referencing others' work — even when you think that the connection is somewhat tenuous.

Rule 3. *Develop your own style.*
No one who has seen Marcel Erné lecture will forget the clever cartoons he draws that succinctly tell the crucial — and sometimes not so crucial — points. For this reason his lectures are always well attended. My own rule of thumb for colloquium talks is that they should be divided into roughly five equal parts; the first part understandable to the man in the street (anyone who has been to junior high school); the second part understandable to the average undergraduate math major; the third part understandable to graduate students in your field and colleagues in other fields; the fourth part understandable to you and specialists in your field; and the last part possibly totally unintelligible to all. This will leave everyone fairly happy — or impressed. You might adopt Marcel's strategy or the above rule of thumb — or someone else's, but let it freely mutate, to become whatever you find gets good results for you. The same goes for "teaching style" or "writing style". Start with what seems to be a good plan and then modify it over time until it "fits" and becomes your own.

Rule 4. *Salesmanship is the key.*
Whether you are trying to get a paper accepted by a journal or trying to get students to accept the ϵ-δ definition of a limit or trying to get a raise in salary, you should spend a bit of preliminary time on a sales strategy. I'll never forget the response to the first paper that I sent off for publication. It was based on results in my dissertation. They were correct and they extended known results of reasonably prestigious mathematicians. So I was quite proud of myself and as soon as I put it in the mail I essentially chalked it up as my first paper. To my horror it was rejected and immediately returned by the journal editor! The reason was that I hadn't "sold" the results. The editor told me that he hadn't forwarded the paper to a referee because I hadn't told him in the introduction why the results in that paper were interesting or important. So even if everything were correct, it wasn't publishable. After I recovered, I thoroughly re-wrote the introduction and included the needed "sales pitch". To my relief it was accepted, and since then

I have kept the need to "sell ideas" always present in the back of my mind. When a student in your class asks "what good is this", the question should not irritate you. Rather, it should alert you to start thinking about and questioning your "sales strategy".

Rule 5. *Eccentricity is really the key.*
Virtually everyone on earth is eccentric. Think about it. For one to be in the center with respect to all attributes is highly unlikely. However, even though this is the case, nearly everyone believes that allowing his or her natural eccentricity to be exposed would be embarrassing — or at least would be highly disadvantageous. As a mathematician one of your key fringe benefits is that all of this is reversed. You are *expected* to be eccentric and this means that (within limits) you can just let your innate weirdness develop naturally. You don't need to stifle yourself. This enhances your own mental health and in many ways adds to each country's GNH (gross national happiness). As an example of this, just think of all the stories that are told and retold about mathematicians. Of course this rule can be used in conjunction with Rule 3. Years ago I realized that my "natural" teaching style should be one of involving students in my classes as much as possible. One time as I came to work I was told by the secretaries that the horticulture club was giving away cantaloupes from the University farm. So I got half a bushel and in my classes that day I tossed cantaloupes to those students who correctly answered my questions. In this same vein I have always tried to give an April fool's lecture when class falls on the right day. I start with mildly wrong statements and proceed until I'm telling whoppers that are so huge that some student finally stops me. (It is often sobering to see how long this can take.) Also along the lines of student involvement, in the last few years I have adopted a "three strikes and you're out" policy in my classes. This means that each time a student discovers a mistake of mine, it counts as a strike (against me). As soon as there are three strikes, the class is over and we all leave, even though there might be much class time remaining. Ironically, this works! Students who are not required to attend class come and carefully monitor for correctness every word that I utter with the hope of getting out of class early. By doing so they learn.

Rule 6. *Don't publish the same result several times.*
It could prove to be embarrassing; see [5].

Rule 10. *You don't really need to obey the above rules.*
Proof: Just remember Rule 1 (and see the third lesson in [4]).

Now some of you who have been paying attention may say that you have been cheated and you haven't received the promised number of rules. To you I say that this whole article has been written using base seven; (i.e., base 10).

References

[1] Gian-Carlo Rota. Ten lessons I wish I had been taught. *Concerns of Young Mathematicians*, **4** No. 25 (1996): http://www.youngmath.org/vol4.25.html#2.

[2] Gian-Carlo Rota. Ten lessons I wish I had been taught. *Indiscrete Thoughts*, Ch. XVIII (pp. 195–203), Birkhäuser Boston (1997). ISBN 3-7643-3866-0

[3] Gian-Carlo Rota. Zehn ungelehrte Lektionen. *Mitt. Dtsch. Math. Ver.*, **2** (1997): 34-38.

[4] Gian-Carlo Rota. Ten lessons I wish I had been taught. *Notices Amer. Math. Soc.*, **44** No. 1 (1997): 22-25.

[5] Paul Nevai. Three lessons I wish I had never been taught by Gian-Carlo Rota. *Notices Amer. Math. Soc.*, **44** No. 3 (1997): 309.

George E. Strecker
Department of Mathematics
Kansas State University
Manhattan, Kansas 66506, USA
strecker@math.ksu.edu

Connections and Polarities

Austin Melton

ABSTRACT Galois connections – also called residuated/residual maps – have been studied and used extensively in both mathematics and computer science. In the 1980s, Galois connections were generalized to connections, and in the 1990s, a counterpart to Galois connections, called Lagois connections, were discovered. Lagois connections, as the name suggests, are similar to Galois connections; however, the "movement" from arbitrary points to image points – which is done via the composite maps of the connections – is in the same "direction" with respect to the order relations in both partially ordered sets. This similarly directed movement seems to be characteristic of (many) computer science applications. In this chapter we present these concepts in their "discovery" order; we also give properties and examples. We begin with a pre-Galois connection concept, called a polarity.

Key words: Polarity, poset system, connection, Galois connection, Lagois connection.

AMS Subject Classifications (2000): 06A15.

1 Introduction

In 1940, G. Birkhoff [1] defined a polarity in terms of a relation between two sets. Using the relation, Birkhoff defined two order-reversing functions between the power sets of the original two sets. These two functions establish "connections" between elements of the two power sets.

In 1944, O. Ore [2] generalized polarities, which establish connections between elements of two power sets (which are partially ordered sets in the subset ordering), to Galois connections which are defined on or between partially ordered sets. Fittingly, Galois theory provides examples of Galois connections.

In 1953, J. Schmidt [3] introduced "order-preserving" Galois connections. These order-preserving Galois connections were also introduced and studied in lattice theory under the name of residuated/residual maps. In computer science, the order-preserving version has proven more useful than the order-reversing version. The reason for this seems to be that in computer science when (partially) equivalent structures are compared, the underlying equivalent structures have isomorphic order relations.

Proofs of most of the propositions are either given in the cited papers or are relatively easy; thus, proofs are not included in this tutorial chapter.

2 Polarities

In 1940, G. Birkhoff [1] defined a polarity as follows:

Definition 2.1 Let ρ be a relation between sets X and Y, i.e.,

$$\rho \subseteq X \times Y.$$

Define two functions

$$F : \mathcal{P}(X) \to \mathcal{P}(Y)$$

and

$$G : \mathcal{P}(Y) \to \mathcal{P}(X)$$

between the power sets of X and Y by

$$F(A) = \{y \in Y | \forall x \in A,\ x\rho y\}$$

and

$$G(B) = \{x \in X | \forall y \in B,\ x\rho y\}.$$

$$(F, \mathcal{P}(X), \rho, \mathcal{P}(Y), G) \text{ or simply } (F, \rho, G)$$

is called a polarity.

Example 2.2 Consider the relation $\leq$ between the integers, **Int**, and the reals, **Re**. For $A \subseteq \mathbf{Int}$,

$$F(A) = \{r \in \mathbf{Re} | a \leq r, \forall a \in A\},$$

which is the set of all real numbers greater than or equal to the largest integer in A if A has a largest integer and is $\emptyset$ if A is nonempty and has no largest element. $GF(A)$ is the set of all integers less than or equal to the largest integer in A if A has a largest integer and is equal to **Int** if A is nonempty and has no largest element. If $A = \emptyset$, then $F(A) = \mathbf{Re}$, and $GF(A) = \emptyset$. For $r \in \mathbf{Re}$, $G(\{r\}) = \{i \in \mathbf{Int} | i \leq r\}$, which is the principal ideal of **Int** generated by the floor of r.

Definition 2.3 Let X and Y be sets, and let $f : X \to Y$ and $g : Y \to X$ be functions. g is said to be a *quasi-inverse* for f if $fgf = f$.

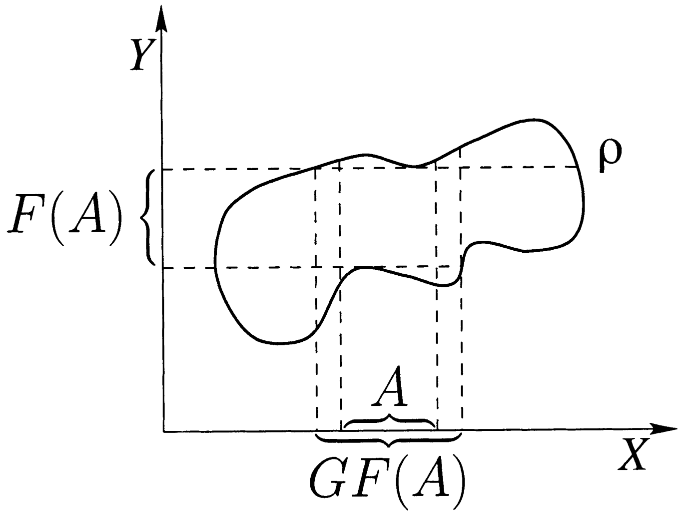


FIGURE 2.1. A polarity $(F, (\mathcal{P}(X), \leq), \rho, (\mathcal{P}(Y), \leq), G)$

Proposition 2.4 *Let $(F, \mathcal{P}(X), \rho, \mathcal{P}(Y), G)$ be a polarity.*

1. *$\mathcal{P}(X)$ and $\mathcal{P}(Y)$ are partially ordered sets, ordered by set inclusion, and F and G are order-reversing functions.*

2. *GF and FG are order-increasing functions, i.e.,*

$$\forall A \subseteq X,\ A \subseteq GF(A)$$

and

$$\forall B \subseteq Y,\ B \subseteq FG(B).$$

3. *G is a quasi-inverse for F, and F is a quasi-inverse for G.*

4. *$(G[\mathcal{P}(Y)], \subseteq)$ and $(F[\mathcal{P}(X)], \subseteq)$ are anti-isomorphic partially ordered sets, with the restrictions of F and G to these sets being anti-isomorphisn Further, these restricted functions are inverses of each other.*

5. *$A \in G[\mathcal{P}(Y)]$ iff $A = GF(A)$ and*

$$B \in F[\mathcal{P}(X)] \text{ iff } B = FG(B).$$

6. *If $B \in F[\mathcal{P}(X)]$, then $G(B)$ is the largest element which is in $\mathcal{P}(X)$ and whose image is B. Likewise, if $A \in G[\mathcal{P}(Y)]$, then $F(A)$ is the largest element which is in $\mathcal{P}(Y)$ and whose image is A. Thus, if*

$$B \in F[\mathcal{P}(X)]$$

and if

$$\mathcal{C} = \{A \in \mathcal{P}(X) | F(A) = B\},$$

then

$$G(B) \in \mathcal{C}, \text{ and for each } A \in \mathcal{C}, A \subseteq G(B).$$

Likewise, if

$$A \in G[\mathcal{P}(Y)]$$

and if

$$\mathcal{D} = \{B \in \mathcal{P}(Y) | G(B) = A\},$$

then

$$F(A) \in \mathcal{D}, \text{ and for each } B \in \mathcal{D}, B \subseteq F(A).$$

Polarities are common and useful. Whenever one has a relation between two sets, then a polarity which naturally relates subsets of the two original sets exists. Further, a polarity is a special case of a Galois connection.

3 Galois Connections

In 1944, O. Ore [2] generalized polarities to Galois connections, which he originally called Galois "connexions".

Definition 3.1 Let $(P, \leq)$ and $(Q, \leq)$ be partially ordered sets, and let $f : P \to Q$ and $g : Q \to P$ be order-reversing functions. $(f, (P, \leq), (Q, \leq), g)$ or simply (f, g) is a Galois connection if

$$\forall p \in P \text{ and } \forall q \in Q, p \leq g(q) \text{ iff } q \leq f(p).$$

There is an alternate definition which is given next.

Definition 3.2 Let $(P, \leq)$ and $(Q, \leq)$ be partially ordered sets, and let $f : P \to Q$ and $g : Q \to P$ be order-reversing functions. $(f, (P, \leq), (Q, \leq), g)$ or simply (f, g) is a Galois connection if

$$\forall p \in P, p \leq gf(p)$$

and

$$\forall q \in Q, q \leq fg(q).$$

Example 3.3 Let $E : F$ be a field extension, and let $\mathcal{P}$ be the set of intermediate fields between F and E ordered by set inclusion. Let $G = G(E, F)$ be the group of all field automorphisms of E that fix F pointwise, and let $\mathcal{Q}$ be the set of subgroups of G also ordered by set inclusion. Define $f : \mathcal{P} \to \mathcal{Q}$ by

$$f(L) = \{g \in G | g(x) = x, \forall x \in L\},$$

and define $g : \mathcal{Q} \to \mathcal{P}$ by

$$g(H) = \{z \in E | \forall h \in H, h(z) = z\}.$$

This is the Galois connection that arises from E. Galois' work that established Galois theory. This example, with more detail, is given in [4].

There is a proposition for Galois connections which is much like Proposition 2.4 for polarities. However, instead of giving this proposition now, we will state it for Galois connections with order-preserving functions.

4 Order-Preserving Galois Connections or Residuated and Residual Mappings

In 1953, J. Schmidt [3] introduced "order-preserving" Galois connections. The definition is, of course, very similar to Definition 3.1. One can think of an order-preserving Galois connection as being a Galois connection as defined earlier but with the order on Q reversed, i.e., with the order relation on Q replaced by its dual order.

Definition 4.1 Let $(P, \leq)$ and $(Q, \leq)$ be partially ordered sets, and let $f : P \to Q$ and $g : Q \to P$ be *order-preserving* maps. $(f, (P, \leq), (Q, \leq), g)$ or simply (f, g) is a(n order-preserving) Galois connection if

$$\forall p \in P \textit{ and } \forall q \in Q, p \leq g(q) \textit{ iff } f(p) \leq q.$$

An interesting difference between order-reversing and order-preserving Galois connections is that, in the case of order-reversing functions, the images $g[Q]$ and $f[P]$ are anti-isomorphic, whereas with order-preserving functions the images $g[Q]$ and $f[P]$ are essentially the same, i.e., they are isomorphic partially ordered sets.

An advantage of order-preserving Galois connections, when compared to order-reversing ones, is that order-preserving Galois connections can be composed and the result is another order-preserving Galois connection.

From now on in this chapter, Galois connections will mean Galois connections with order-preserving functions unless order-reversing functions are explicitly mentioned.

These order-preserving Galois connections were also introduced and studied in lattice theory under the name of residuated and residual maps. In Definition 4.1, f is the residuated map, and g is the residual one. The study of residuated and residual maps is *residuation theory.* A classical reference is *Residuation Theory* by T.S. Blyth and M.F. Janowitz [5].

In computer science, order-preserving Galois connections have proven more useful and common than order-reversing ones. The reason for this seems to be that in computer science, when (partially) equivalent structures are compared, the underlying structures (i.e., the $g[Q]$ and $f[P]$) are essentially the same – that is, these structures are order isomorphic. For examples of Galois connections in computer science, see [6].

Definition 4.2 Let $(P, \leq)$ and $(Q, \leq)$ be partially ordered sets, and $f : P \to Q$ and $g : Q \to P$ be order-preserving functions. Then $(f, (P, \leq), (Q, \leq), g)$ or simply (f, g) is called a poset system.

Definition 4.3 Let $(f, (P, \leq), (Q, \leq), g)$ be a poset system. An element $p \in P$ or $q \in Q$ is said to be a fixed point of the poset system, or simply a fixed point, if $p = gf(p)$ or $q = fg(q)$, respectively.

Proposition 4.4 *Let $(f, (P, \leq), (Q, \leq), g)$ be a Galois connection.*

1. *gf is an increasing function, i.e.,*

$$\forall p \in P,\ p \leq gf(p),$$

and fg is a decreasing function, i.e.,

$$\forall q \in Q,\ fg(q) \leq q.$$

2. *f and g uniquely determine each other.*

3. *f and g are quasi-inverses of each other.*

4. *$p \in P$ is a fixed point iff $p \in g[Q]$ and $q \in Q$ is a fixed point iff $q \in f[P]$.*

5. *f is injective iff g is surjective iff $gf = id_P$ and g is injective iff f is surjective iff $fg = id_Q$.*

6. *$g[Q]$ and $f[P]$ are isomorphic partially ordered sets, with the restrictions*

$$f\Big|_{g[Q]}^{f[P]} : g[Q] \to f[P],$$

$$g\Big|_{f[P]}^{g[Q]} : f[P] \to g[Q]$$

being order isomorphisms, and these two restricted functions are inverses to each other.

7. *f preserves joins, and g preserves meets.*

8. *If P and Q are complete lattices, then so are $g[Q]$ and $f[P]$, though they may not be sublattices.*

9. *If*

$$q \in f[P] \text{ and if } R = \{p \in P | f(p) = q\},$$

then

$$g(q) \in R, \text{ and for each } p \in R, p \leq g(q).$$

Similarly, if

$$p \in g[Q] \text{ and if } S = \{q \in Q | g(q) = p\},$$

then

$$f(p) \in S, \text{ and for each } q \in S, \ f(p) \leq q.$$

Remark 4.5 As there was an alternative definition for order-reversing Galois connections, there is an alternative definition for Galois connections.

Definition 4.6 Let $(P, \leq)$ and $(Q, \leq)$ be partially ordered sets, and let $f : P \to Q$ and $g : Q \to P$ be **order-preserving** maps. $(f, (P, \leq), (Q, \leq), g)$ or simply (f, g) is a(n order-preserving) Galois connection if

$$\forall p \in P, p \leq gf(p)$$

and

$$\forall q \in Q, fg(q) \leq q.$$

Remark 4.7 There is a natural way to use relations to get an order-preserving counterpart to polarities and order-reversing Galois connections. We begin by letting ρ be a relation between sets X and Y, i.e., $\rho \subseteq X \times Y$, and by defining a function

$$\rho^* : \mathcal{P}(X) \to \mathcal{P}(Y)$$

by

$$\rho^*(A) = \{y \in Y | \exists a \in A, a\rho y\}.$$

Example 4.8 If $\rho \subseteq \mathbf{Re} \times \mathbf{Re}$ is defined by $x\rho y$ if and only if $x \geq 0$ and $x = |y|$, then

$$\rho^*(\{5,6\}) = \{-6,-5,5,6\},$$

$$\rho^*(\{-5,6\}) = \{-6,6\},$$

and

$$\rho^*(\{-5,-6\}) = \emptyset.$$

Definition 4.9 Let ρ be a relation between sets X and Y, i.e.,

$$\rho \subseteq X \times Y.$$

Define two functions

$$F : \mathcal{P}(X) \to \mathcal{P}(Y)$$

and

$$G : \mathcal{P}(Y) \to \mathcal{P}(X)$$

between the power sets of X and Y by

$$F(A) = \rho^*(A)$$

and

$$G(B) = \{x \in X | \rho^*(\{x\}) \subseteq B\}.$$

$(F, \mathcal{P}(X), \rho, \mathcal{P}(Y), G)$ or simply (F, ρ, G) is an order-preserving polarity, i.e., $(F, \mathcal{P}(X), \mathcal{P}(Y), G)$ is a Galois connection.

5 Connections

In 1982, H. Crapo [8] generalized Galois connections to connections.

Definition 5.1 Let $(P, \leq)$ and $(Q, \leq)$ be partially ordered sets, and $f : P \to Q$ and $g : Q \to P$ be order-preserving functions, then $(f, (P, \leq), (Q, \leq), g)$ or simply (f, g) is called a connection if g is a quasi-inverse for f and f is a quasi-inverse for g.

Example 5.2 Let **Re** be the set of reals; let **Int** be the set of integers; and define $f : \mathbf{Re} \to \mathbf{Int}$ and $g : \mathbf{Int} \to \mathbf{Re}$ so that f is the "rounding" function and g is the identity. Thus, $f(r) =$ the ceiling of r if the standard decimal representation of r is greater than or equal to 0.5, and $f(r) =$ the floor of r if the standard decimal representation of r is less than 0.5. $g(i) = i$. (f, g) and (g, f) are connections.

Whenever $f : X \to Y$ is a function from set X to set Y, f defines a partition on X. The equivalence classes of the partition are the fibers of f, i.e., nonempty sets of the form $f^{-1}(y)$ for $y \in Y$. We will denote this partition by $\mathcal{PART}(f)$ or by $\mathcal{PART}(X)$ when the corresponding function is clearly known. We may denote an element of $\mathcal{PART}(X)$ by X_i or by $[x]$ for $x \in X$. In the case of poset systems, the $\mathcal{PART}(P)$ and $\mathcal{PART}(Q)$ are not especially interesting; however, they start to become interesting for connections.

Proposition 5.3 *If $((P, \leq), f, g, (Q, \leq))$ is a connection, then $\mathcal{PART}(P)$ and $\mathcal{PART}(Q)$ are partially ordered sets defined by*

$$\begin{gathered} \text{for } P_1, P_2 \in \mathcal{PART}(P), \\ P_1 \leq P_2 \\ \text{if and only if} \\ f(p_1) \leq f(p_2) \text{ for any } p_1 \in P_1, p_2 \in P_2 \end{gathered}$$

and

$$\begin{gathered} \text{for } Q_1, Q_2 \in \mathcal{PART}(Q), \\ Q_1 \leq Q_2 \\ \text{if and only if} \\ g(q_1) \leq g(q_2) \text{ for any } q_1 \in Q_1, q_2 \in Q_2. \end{gathered}$$

Proposition 5.4 *Let $((P, \leq), f, g, (Q, \leq))$ be a connection.*

1. *The fixed points of P are precisely the elements of $g[Q]$ and the fixed points of Q are precisely the elements of $f[P]$.*

2. *The set $g[Q]$ and $f[P]$ are isomorphic partially ordered sets, and the restrictions*

$$f\Big|_{g[Q]}^{f[P]} : g[Q] \to f[P],$$

$$g\Big|_{f[P]}^{g[Q]} : f[P] \to g[Q]$$

are partial-order isomorphisms.

3. *The partially ordered sets $\mathcal{PART}(P)$ and $\mathcal{PART}(Q)$ are isomorphic with isomorphisms being*

$$f^* : \mathcal{PART}(P) \to \mathcal{PART}(Q)$$
defined by
$$f^*([p]) = [f(p)].$$

and

$$g^* : \mathcal{PART}(Q) \to \mathcal{PART}(P)$$
defined by
$$g^*([q]) = [g(q)].$$

6 Lagois Connections

If one looks at the computer science Galois connections in [6], one notices a common characteristic – in each example, one of the functions is injective. These are not examples of general Galois connections. While writing [6] and for a time thereafter, we (the authors of that paper – Melton, Schmidt, and Strecker) tried to find more general examples of Galois connections in computer science; in particular, we were trying to generalize the examples in that paper to Galois connections in which neither map would be injective. We did not succeed.

In a Galois connection, the function gf is order-increasing and the function fg is order-decreasing. Thus, when one takes an arbitrary point in P and when one "moves" or transforms this point via gf to a fixed point in P, then movement is *up* in the order relation in P. However, when one takes an arbitrary point in Q and when one "moves" it to a fixed point via fg in Q, the movement is *down* in the order relation on Q. However, in computer science when one is comparing potentially equivalent or similar partially ordered sets, it seems that the movement to fixed points is in the same direction in both sets. This is true for the examples in [7] which generalize two of the examples in [6]. However, the generalized examples in [7] are not Galois connections; they are Lagois connections. It was in trying to develop a mathematical construction that would model an example like the complier equivalence example in [7] that Lagois connections were discovered.

Definition 6.1 Let $(f, (P, \leq), (Q, \leq), g)$ be a connection. (f, g) is a Lagois connection if

$$gf \text{ and } fg \text{ are both increasing functions}$$

and if

$$f \text{ and } g \text{ are quasi-inverses of each other.}$$

One could define two types of Lagois connections, increasing and decreasing ones; that is, instead of requiring that gf and fg be increasing functions, one could require that both gf and fg be decreasing functions.

Example 6.2 Let (X, τ) be a topological space. Define

$$f : (\mathcal{P}(X), \subseteq) \to (\mathcal{P}(X), \supseteq)$$

by

$$f(A) = X - cl(A)$$

and define

$$g : (\mathcal{P}(X), \supseteq) \to (\mathcal{P}(X), \subseteq)$$

by

$$g(A) = cl(X - A),$$

where $cl(A)$ is the topological closure of the set A in the topology on X.

$(f, (\mathcal{P}(X), \subseteq), (\mathcal{P}(X), \supseteq), g)$ is a Lagois connection, and gf is the topological closure operator while fg is the topological interior operator. [7]

Proposition 6.3 *Let $(f, (P, \leq), (Q, \leq), g)$ be a Lagois connection. (Cf. [7])*

1. *f and g uniquely determine each other.*
2. *$p \in P$ is a fixed point iff $p \in g[Q]$ and $q \in Q$ is a fixed point iff $q \in f[P]$.*
3. *f is injective iff g is surjective iff $gf = id_P$ and g is injective iff f is surjective iff $fg = id_Q$.*
4. *$g[Q]$ and $f[P]$ are isomorphic partially ordered sets, with the restrictions*

$$f \Big|_{g[Q]}^{f[P]} : g[Q] \to f[P],$$

$$g \Big|_{f[P]}^{g[Q]} : f[P] \to g[Q]$$

being order isomorphisms, and these two restricted functions are inverses to each other.

5. *If P and Q are complete lattices, then so are $g[Q]$ and $f[P]$, though they may not be sublattices.*

6. *If*

$$q \in f[P] \text{ and if } R = \{p \in P | f(p) = q\},$$

then

$$g(q) \in R, \text{ and for each } p \in R, \ p \leq g(q).$$

Similarly, if

$$p \in g[Q] \text{ and if } S = \{q \in Q | g(q) = p\},$$

then

$$f(p) \in S, \text{ and for each } q \in S, \ q \leq f(p).$$

There is no known (natural) method for constructing Lagois connections from relations.

References

[1] G. Birkhoff. *Lattice Theory*, 1st ed., Amer. Math. Soc., Colloquium publications, v. 25, New York, 1940. (3rd ed, 1967, Providence, RI)

[2] O. Ore. Galois connexions, *Trans. Amer. Math. Soc.*, **55** (1944), 493–513.

[3] J. Schmidt. Beiträge zur Filtertheorie, II, *Math. Nachr.*, **10** (1953), 197–232.

[4] M. Erné, J. Koslowski, A. Melton, G.E. Strecker. A Primer on Galois Connections, *Papers on General Topology and Applications*, *Annals of the New York Academy of Sciences*, **704** (1993), 103–125.

[5] T.S. Blyth, M.F. Janowitz. *Residuation Theory*, 1st ed., Pergamon Press, Oxford, 1972.

[6] A. Melton, D.A. Schmidt, G.E. Strecker. Galois Connections and Computer Science Applications, *Category Theory and Computer Programming*, *Lecture Notes in Computer Science*, Springer-Verlag, **704** (1986), 299–312.

[7] A. Melton, B.S.W. Schröder, G.E. Strecker. Lagois connections – a counterpart to Galois connections, *Theoretical Computer Science*, **136** (1994), 79–107.

[8] H. Crapo. Ordered sets: retracts and connections, *J. of Pure and Applied Algebra*, **23** (1982), 13–28.

Austin Melton
Department of Mathematics and Computer Science
Kent State University
Kent, Ohio 44242, USA
melton@mcs.kent.edu

Categorical Closure Operators

G. Castellini[1]

ABSTRACT A brief survey of the development of the theory of closure operators is presented. Results concerning the applications of the theory to epimorphisms, separation, compactness and connectedness are also included together with a number of supporting examples.

Key words: Closure operator, Galois connection, factorization structure, epimorphism, separation, compactness, connectedness, constant morphism.

AMS Subject Classifications (2000): 18-02.

0 Introduction

Undoubtedly, the inspiring work for the theory of categorical closure operators was Salbany's paper [S]. In this paper a particular closure construction in the category **Top** of topological spaces was introduced. This construction was then extended by other authors to an arbitrary category $\mathcal{X}$ and in an indirect way it led to the general concept of categorical closure operator. The first one to see in Salbany's closure construction a great potential for further development was Eraldo Giuli who in [G_1] used it to obtain a characterization of the epimorphisms in epireflective subcategories of **Top**. The first paper to present a formal introduction of the above operator in **Top** was [DG_1]. This was followed by [DG_2] and [GH_1] in which a diagonal theorem for quotient reflective subcategories of **Top** was proved and some questions about co-well poweredness of epireflective subcategories of **Top** were answered.

The first attempt to introduce a general notion of closure operator in a concrete category was made by Castellini in [C_1]. In this paper an extended version of Salbany's construction was used to study the surjectivity of epimorphisms in several subcategories of abelian groups. Moreover, a dual notion was used to study the monomorphisms. Finally in [DG_3] the current notion of categorical closure operator was introduced in a category

[1]The author acknowledges support from the Research Office of the Faculty of Arts and Sciences of the University of Puerto Rico – Mayagüez campus.

$\mathcal{X}$ together with some important basic properties. This paper laid the basis for further development of the theory. For instance [CS], [GMT] and [T_1] dealt with the diagonal theorem in an arbitrary category, among other topics. In [K] some of the results in [DG_3] were sharpened. Among the papers that have dealt with the general theory of categorical closure operators we would like to mention [DGT], [CKS_{1-5}], [CG_{1-2}], [DGTo], [F_3] and [Cl_2].

The basic idea behind the theory of categorical closure operators is to provide a tool that allows us to extend to an arbitrary category $\mathcal{X}$ all those concepts that in **Top** were defined in terms of the classical Kuratowski operator. This has made possible the introduction of topological concepts in categories without any topology at all. Among these concepts I would like to mention separation, compactness and connectedness. Relative to the aspect of applications of the general theory I would like to mention the following papers: [CS], [CGT_1], [Cl_3], [CT_1], [GMT] for separation, [C_{3-4}], [CGT_1], [DG_{4-6}], [F_{1-2}], [FJ], [FS], [FW_{1-6}] [G_3], [GH_2] for compactness, [AC_2], [CC_3], [C_{5-8}], [CH], [Cl_5], [CT_1], [P] for connectedness.

We conclude with a list of papers that have either contributed to the development of the general theory of categorical closure operators or that have provided applications of the theory to specific categories: [AC_1], [C_2], [CC_{1-2}], [CGT_2], [Cl_1], [Cl_4], [CT_{2-3}], [D], [DG_{7-8}], [DW], [Fe_{1-3}], [G_{2-3}], [GT], [GS], [Ho_2], [L], [So_{1-2}], [St].

The aim of this paper is to introduce the reader to the theory of categorical closure operators. In Section 1 the general theory will be presented. The proofs of the results have not been included. However, the reader interested in a deeper understanding of the topic can find them in the appropriate referenced papers, in [Ho_1] or in [DT]. In Sections 2–5, applications to some specific concepts will be presented. Precisely, Section 2 deals briefly with Epimorphisms, Section 3 refers to Separation and Sections 4 and 5 include the main results on Compactness and Connectedness, respectively. We would like to observe in this respect that only those papers that have dealt with the above topics from the point of view of categorical closure operators have been considered here.

We would like to emphasize that in order to keep this article within a certain number of pages, we have only included those results that have had a major impact either on the development of the general theory or on the applications. However, every single paper we included in the references contains much more material than what we could report here. Therefore we strongly encourage the reader seeking more details to consult the original works.

We have tried very hard to reconstruct the development of the theory of categorical closure operators to the best of our knowledge. However, as in any attempt to write a survey article, it is difficult to be aware of all the papers published on a given subject. Therefore, it is likely that we may have missed some related publications. In the event that this may have occurred or that we may have wrongly attributed some ideas, we would like to make clear that it was absolutely unintentional and we apologize beforehand for that.

We use the terminology of [AHS] throughout the paper.

1 The General Theory

The aim of this section is to introduce the current notion of categorical closure operator (which, as customary, will be simply called closure operator), together with the most commonly used results and some examples. We would like to observe that the first attempt at defining this new notion of closure operator started in the category **Top** of topological spaces ([DG_1]) and subsequently in a concrete category $U : \mathcal{A} \longrightarrow \mathcal{X}$ ([C_1]). However, in what follows we will skip these two preliminary steps and go straight to the well established definition that is basically a product of [DG_3].

Throughout this section we consider a category $\mathcal{X}$ and a fixed class $\mathcal{M}$ of $\mathcal{X}$-monomorphisms, which contains all $\mathcal{X}$-isomorphisms. It is assumed that $\mathcal{X}$ is $\mathcal{M}$-complete; i.e.:

(1) $\mathcal{M}$ is closed under composition;

(2) Pullbacks of $\mathcal{M}$-morphisms exist and belong to $\mathcal{M}$, and multiple pullbacks of (possibly large) families of $\mathcal{M}$-morphisms with common codomain exist and belong to $\mathcal{M}$.

One of the consequences of the above assumptions is that there is a uniquely determined class $\mathbf{E}$ of sinks in $\mathcal{X}$ such that $\mathcal{X}$ is an $(\mathbf{E}, \mathcal{M})$-*category for sinks*, that is:

(a) each of $\mathbf{E}$ and $\mathcal{M}$ is closed under compositions with isomorphisms;

(b) $\mathcal{X}$ has $(\mathbf{E}, \mathcal{M})$-factorizations (of sinks); i.e., each sink $\mathbf{s}$ in $\mathcal{X}$ has a factorization $\mathbf{s} = m \circ \mathbf{e}$ with $\mathbf{e} \in \mathbf{E}$ and $m \in \mathcal{M}$, and

(c) $\mathcal{X}$ has the unique $(\mathbf{E}, \mathcal{M})$-diagonalization property; i.e., if $B \xrightarrow{g} D$ and $C \xrightarrow{m} D$ are $\mathcal{X}$-morphisms with $m \in \mathcal{M}$, and $\mathbf{e} = (A_i \xrightarrow{e_i} B)_I$ and $\mathbf{s} = (A_i \xrightarrow{s_i} C)_I$ are sinks in $\mathcal{X}$ with $\mathbf{e} \in \mathbf{E}$, such that $m \circ \mathbf{s} = g \circ \mathbf{e}$, then there exists a unique diagonal $B \xrightarrow{d} C$ such that for every $i \in I$ the following diagrams commute:

$$\begin{tikzcd} A_1 \arrow[r, "e_i"] \arrow[d, "s"'] & B \arrow[dl] \\ C & \end{tikzcd} \qquad \text{and} \qquad \begin{tikzcd} & B \arrow[dl, "d"'] \arrow[d, "g"] \\ C \arrow[r, "m"'] & D \end{tikzcd}$$

That $\mathcal{X}$ is an $(\mathbf{E}, \mathcal{M})$-category implies the following features of $\mathcal{M}$ and $\mathbf{E}$ (cf. [AHS] for the dual case):

Proposition 1.1

(1) *Every isomorphism is in both $\mathcal{M}$ and* $\mathbf{E}$ *(as a singleton sink).*

(2) $\mathcal{M}$ *is closed under $\mathcal{M}$-relative first factors, i.e., if $n \circ m \in \mathcal{M}$, and $n \in \mathcal{M}$, then $m \in \mathcal{M}$.*

(3) $\mathcal{M}$ *is closed under composition.*

(4) *Pullbacks of $\mathcal{X}$-morphisms in $\mathcal{M}$ exist and belong to $\mathcal{M}$.*

(5) *The $\mathcal{M}$-subobjects of every $\mathcal{X}$-object form a (possibly large) complete lattice; suprema are formed via $(\mathbf{E}, \mathcal{M})$-factorizations and infima are formed via intersections.*

Clearly any $(\mathbf{E}, \mathcal{M})$-category for sinks is also an $(\mathcal{E}, \mathcal{M})$-category for single morphisms, with $\mathcal{E}$ consisting of all morphisms (singleton sinks) belonging to $\mathbf{E}$. Notice that in the above proposition, the word "lattice" is to be understood in a generalized sense for not necessarily antisymmetric pre-orders. Moreover, throughout the paper we will use the expression $\mathcal{M}$-subobject for both $m \in \mathcal{M}$ and the corresponding equivalence class of elements of $\mathcal{M}$ and for $X \in \mathcal{X}$, the symbol $\mathcal{M}_X$ will be often used to denote all $\mathcal{M}$-subobjects of X.

If $X \xrightarrow{f} Y$ is an $\mathcal{X}$-morphism and $M \xrightarrow{m} X$ is an $\mathcal{M}$-subobject, then $M \xrightarrow{e_{f\circ m}} M_f \xrightarrow{m_f} Y$ will denote the $(\mathbf{E}, \mathcal{M})$-factorization of $f \circ m$. $M_f \xrightarrow{m_f} Y$ will be called the direct image of m along f. If $N \xrightarrow{n} Y$ is an $\mathcal{M}$-subobject, then the pullback $f^{-1}(N) \xrightarrow{f^{-1}(n)} X$ of n along f will be called the inverse image of n along f. Whenever no confusion is likely to arise, to simplify the notation we will denote the morphism $e_{f\circ m}$ simply e_f.

Definition 1.2 A *closure operator* C on $\mathcal{X}$ (with respect to $\mathcal{M}$) is a family $\{(\ )^C_X\}_{X \in \mathcal{X}}$ of functions on the $\mathcal{M}$-subobject lattices of $\mathcal{X}$ with the following properties that hold for each $X \in \mathcal{X}$:

(a) [expansiveness] $m \le (m)^C_X$, for every $\mathcal{M}$-subobject $M \xrightarrow{m} X$;

(b) [order-preservation] $m \le n \Rightarrow (m)^C_X \le (n)^C_X$ for every pair of $\mathcal{M}$-subobjects of X;

(c) [morphism-consistency] If p is the pullback of the $\mathcal{M}$-subobject $M \xrightarrow{m} Y$ along some $\mathcal{X}$-morphism $X \xrightarrow{f} Y$ and q is the pullback of $(m)^C_Y$ along f, then $(p)^C_X \le q$, i.e., the closure of the inverse image of m is less than or equal to the inverse image of the closure of m.

Condition (a) implies that for every closure operator C on $\mathcal{X}$, every $\mathcal{M}$-subobject $M \xrightarrow{m} X$ has a canonical factorization

$$\begin{array}{ccc} M & \xrightarrow{t} & (M)^C_X \\ & m \searrow & \downarrow (m)^C_X \\ & & X \end{array}$$

where $((M)^C_X, (m)^C_X)$ is called the *C-closure* of the subobject (M, m).

When no confusion is likely we will write m^C rather than $(m)^C_X$ and for notational symmetry we will denote the morphism t by m_C. Notice that an alternative notation to $(m)^C_X$ that has been extensively used in the literature is $C_X(m)$. We will use the first type of notation nearly throughout the paper. However, there are some instances in Sections 4 and 5 where the second notation will be used. This is done with the purpose of making the reader wishing to consult the original papers more at ease with the notation being used there.

Proposition 1.3 *Under condition (b), the morphism-consistency condition (c) is equivalent to the following statement concerning direct images: if $M \xrightarrow{m} X$ is an $\mathcal{M}$-subobject and $X \xrightarrow{f} Y$ is a morphism, then $((m)^C_Y)_f \leq (m_f)^C_Y$, i.e., the direct image of the closure of m is less than or equal to the closure of the direct image of m.*

Proposition 1.4 *Under condition (a), both order-preservation and morphism-consistency, i.e., conditions (b) and (c) together are equivalent to the following: given (M, m) and (N, n) $\mathcal{M}$-subobjects of X and Y, respectively, if f and g are morphisms such that $n \circ g = f \circ m$, then there exists a unique morphism d such that the following diagram*

$$\begin{tikzcd}
M \arrow[rr, "g"] \arrow[dr, "m_C"] \arrow[dd, "m"'] & & N \arrow[dr, "n_C"] \arrow[dd, "n" near start] & \\
& M^C \arrow[rr, "d" near start] \arrow[dl, "m^C"] & & N^C \arrow[dl, "n^C"] \\
X \arrow[rr, "f"'] & & Y &
\end{tikzcd}$$

commutes.

Remark 1.5 If we regard $\mathcal{M}$ as a full subcategory of the arrow category of $\mathcal{X}$, with the codomain functor from $\mathcal{M}$ to $\mathcal{X}$ denoted by U, then the above definition can also be stated in the following way: A *closure operator* on $\mathcal{X}$ (with respect to $\mathcal{M}$) is a pair $C = (\gamma, F)$, where F is an endofunctor on $\mathcal{M}$ that satisfies $UF = U$, and γ is a natural transformation from $id_{\mathcal{M}}$ to F that satisfies $(id_U)\gamma = id_U$ (cf. [DG3]).

In any category, the operators that for any $M \xrightarrow{m} X$ are defined by $M^{C} = M$ and $M^{C} = X$ always form a closure operator normally called the discrete and the indiscrete closure, respectively. In what follows we provide a list of more interesting examples.

Example 1.6 Consider the category **Top** of topological spaces with the (*episink*, *embedding*)-factorization structure and let $M \xrightarrow{m} X$ be an embedding. The following are examples of closure operators on **Top**.

(a). The usual topological closure of M, that is the intersection of all closed subsets of X containing M.

(b). The intersection of all clopen subsets of X containing M.

(c). The union of M with all connected subsets of X which intersect M.

(d). The b-closure of M ($b(M)$) that consists of all $x \in X$ such that for every neighborhood U of x, $M \cap Cl(x) \cap U \neq \emptyset$, where $Cl(x)$ denotes the topological closure of the subset $\{x\}$ (cf. [B], [NW]).

(e). $M^{C} = \{y \in X : \exists x \in M \text{ with } Cl(x) = Cl(y)\}$. If $X \xrightarrow{r_0} r_0X$ is the $\mathbf{Top_0}$-reflection, then $M^{C} = r_0^{-1}r_0(M)$. Moreover, $M^{C} \subseteq b(M)$.

(f). The θ-closure of M ($\theta(M)$) that consists of all $x \in X$ such that for every neighborhood U of x, $M \cap Cl(U) \cap U \neq \emptyset$, where $Cl(U)$ denotes the topological closure of the subset U (cf. [V]).

(g). The sequential closure of M ($\Sigma(M)$) that consists of all points $x \in X$ such that there is a sequence in M converging to x.

(h). The union of M with all indiscrete subsets of X which intersect M.

(k). The union of M with all absolutely connected subsets of X that intersect M. We recall that a topological space X is absolutely connected if it cannot be decomposed into any disjoint family $\mathcal{L}$ of nonempty closed subsets with $|\mathcal{L}| > 1$.

Example 1.7 Consider the category **Grp** of groups with the (*episink*, *monomorphism*)-factorization structure and let $M \xrightarrow{m} X$ be a subgroup. The following are examples of closure operators on **Grp**.

(a). The normal closure of M, that is the intersection of all normal subgroups of X containing M.

(b). The intersection of all normal subgroups K of X containing M such that X/K is abelian.

(c). The subgroups generated by M and by all perfect subgroups of X. We recall that a group is perfect if it agrees with the subgroup generated by its commutators.

(d). The subgroup generated by M and by all simple subgroups of X. We recall that a group is simple if it does not have any non-trivial proper normal subgroup.

Example 1.8 Consider the category **Ab** of abelian groups with the (*episink*, *monomorphism*)-factorization structure and let $M \xrightarrow{m} X$ be a subgroup. The following are examples of closure operators on **Ab**.

(a). The operator $M^{C} = M + tor(X)$ where $tor(X)$ denotes the torsion subgroup of X.

(b). The operator $M^{C} = M + div(X)$ where $div(X)$ denotes the largest divisible subgroup of X.

(c). For any epireflective subcategory $\mathcal{A}$, the operator $M^{C} = M + Ker(r_X)$, where $X \xrightarrow{r_X} rX$ is the $\mathcal{A}$-reflection of X.

(d). The intersection of all subgroups K of X containing M such that X/K is torsion-free.

(e). The intersection of all subgroups K of X containing M such that X/K is reduced. We recall that an abelian group is reduced if it does not have any non-trivial divisible subgroup.

(f). For any torsion theory $(\mathcal{T}, \mathcal{F})$, the subgroup generated by M and by all subgroups of X that belong to $\mathcal{T}$.

Definition 1.9 Given a closure operator C, we say that $m \in \mathcal{M}$ is *C-closed* if m_C is an isomorphism. An $\mathcal{X}$-morphism f is called *C-dense* if for some (and hence every) $(\mathbf{E}, \mathcal{M})$-factorization (e, m) of f we have that m^{C} is an isomorphism.

As already observed, for every $\mathcal{M}$-subobject m one obtains a factorization $m = m^{C} \circ m_C$. Contrary to what happens to Kuratowski operators in topology, in general it is not true that m factors via a C-closed and a C-dense morphism. To obtain a result in this direction we need to introduce the following concepts. We call C *idempotent* provided that m^{C} is C-closed for every $m \in \mathcal{M}$. C is called *weakly hereditary* if m_C is C-dense for every $m \in \mathcal{M}$.

Examples of idempotent closure operators are (a–b), (d–e) of 1.6, (a–b) of 1.7 and (a–e) of 1.8. Examples of weakly hereditary closure operators are (a), (c–d), (g–k) of 1.6, (c–d) of 1.7 and (f) of 1.8. However, for instance in 1.6, (f) is neither idempotent nor weakly hereditary, (b) is idempotent but not weakly hereditary and (g) is weakly hereditary but not idempotent.

We denote the conglomerate of all closure operators on $\mathcal{M}$ by $CL(\mathcal{X}, \mathcal{M})$ pre-ordered as follows: $C \sqsubseteq D$ if $m^{C} \leq m^{D}$ for all $m \in \mathcal{M}$ (where $\leq$ is the usual order on subobjects). Notice that arbitrary suprema and infima exist in $CL(\mathcal{X}, \mathcal{M})$; they are formed pointwise in the $\mathcal{M}$-subobject fibers. We will also use $iCL(\mathcal{X}, \mathcal{M})$ for the restriction to the idempotent ones, and $wCL(\mathcal{X}, \mathcal{M})$ for the weakly hereditary ones.

Definition 1.10 For pre-ordered conglomerates $\mathcal{X} = (\mathbf{X}, \sqsubseteq)$ and $\mathcal{Y} = (\mathbf{Y}, \sqsubseteq)$, a *Galois connection* $\mathcal{X} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{Y}$ consists of order preserving functions F and G that satisfy $F \dashv G$, i.e., $x \sqsubseteq GF(x)$ for every $x \in \mathbf{X}$ and

$FG(y) \sqsubseteq y$ for every $y \in \mathbf{Y}$. (G is adjoint and has F as coadjoint).

If $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ are such that $F(x) = y$ and $G(y) = x$, then x and y are said to be corresponding fixed points of the Galois connection $(\mathcal{X}, F, G, \mathcal{Y})$ (we may use at times the shorter notation (F, G)). To be more precise, we may sometimes make use of the expressions "left fixed point" and "right fixed point" for x and y, respectively.

In order to adjust ourselves to different types of notations that have been used in the literature, we will use at times the simpler notation $\mathcal{X} \xrightarrow{F} \mathcal{Y}$ to denote a Galois connection from $\mathcal{X}$ to $\mathcal{Y}$. In this case, F stands for the pair $\langle F_*, F^* \rangle$ and it is understood that the functions $\mathcal{X} \xrightarrow{F_*} \mathcal{Y}$ and $\mathcal{Y} \xrightarrow{F^*} \mathcal{X}$ form respectively the coadjoint and the adjoint part of the Galois connection F.

Properties and many examples of Galois connections can be found in [EKMS].

For a class A we let $P(A)$ denote the conglomerate of all subclasses of A, partially ordered by inclusion. We recall that any relation $\mathcal{R}$ between classes A and B, i.e., $\mathcal{R} \subseteq A \times B$ induces a Galois connection $P(A) \xrightarrow{\phi} P(B)^{\mathbf{op}}$, called a *polarity*, whose adjoint and coadjoint parts are given by

$$\phi^*(V) = \{a \in A : \forall\, b \in V, (a, b) \in \mathcal{R}\} \qquad \text{for } V \subseteq B,$$

$$\phi_*(U) = \{b \in B : \forall\, a \in U, (a, b) \in \mathcal{R}\} \qquad \text{for } U \subseteq A.$$

Let $\mathcal{M}^C$ ($\mathcal{E}^C$) denote the class of all C-closed $\mathcal{M}$-subobjects (C-dense morphisms). The following results can be found mostly in [DG3] and [DGT].

Proposition 1.11 *$\mathcal{X}$ has the $(\mathcal{E}^C, \mathcal{M}^C)$-diagonalization property.*

Proposition 1.12 *For an idempotent closure operator C the following are equivalent:*

(a) C is weakly hereditary;

(b) $\mathcal{M}^C$ is closed under composition;

(c) $\mathcal{X}$ has $(\mathcal{E}^C, \mathcal{M}^C)$-factorizations.

Corollary 1.13 *For an idempotent and weakly hereditary closure operator C one has that $\mathcal{E}^C$ and $\mathcal{M}^C$ are closed under composition in $\mathcal{X}$, $\mathcal{M}^C$ is closed under the formation of pullbacks and limits in $\mathcal{X}$ and $\mathcal{E}^C$ has the dual properties.*

Theorem 1.14 *There is a Galois equivalence between the conglomerate of all subclasses of $\mathcal{M}$ which are part of a factorization system and the conglomerate of all idempotent and weakly hereditary closure operators on $\mathcal{X}$ with respect to $\mathcal{M}$. In both conglomerates, arbitrary infima and suprema exist.*

Given two $\mathcal{M}$-subobjects $M \xrightarrow{m} X$ and $N \xrightarrow{n} X$ such that $m \leq n$, then m_n will denote the morphism from M to N such that $n \circ m_n = m$.

Definition 1.15

(a) C is called *additive* if for every $X \in \mathcal{X}$ and m, n $\mathcal{M}$-subobjects of X, $m^{^C} \vee n^{^C} \simeq (m \vee n)^{^C}$.

(b) Let $\mathcal{M}' \subseteq \mathcal{M}$. Then C is said to be *hereditary with respect to* $\mathcal{M}'$ if whenever m, n are $\mathcal{M}$-subobjects with $m \leq n$ and $n \in \mathcal{M}'$, we have that $n \circ m_n^{^C} \simeq n \wedge m^{^C}$. If $\mathcal{M}' = \mathcal{M}$, then C is simply called *hereditary.*

For instance (a) and (d) of 1.6 are hereditary closure operators, (a–b), (d), (f–g) are additive but (f) is not hereditary. Non-additivity is harder to detect. Examples can be found in [DG$_1$] and [CC$_1$].

Let $hCL(\mathcal{X}, \mathcal{M})$ and $aCL(\mathcal{X}, \mathcal{M})$ denote the conglomerates of all hereditary and all additive closure operators on $\mathcal{M}$, respectively, with the order inherited from $CL(\mathcal{X}, \mathcal{M})$.

We have the following:

Proposition 1.16 *Let C be idempotent. Then:*

(1) C is weakly hereditary if and only if $n \circ m_n^{^C} \simeq n \wedge m^{^C}$ holds whenever $m \leq n$ and n is C-closed;

(2) C is hereditary if and only if it is weakly hereditary and $n \circ m_n^{^C} \simeq n \wedge m^{^C}$ holds whenever $m \leq n$ and n is C-dense;

(3) C is additive if and only if for every $X \in \mathcal{X}$, the class of C-closed $\mathcal{M}$-subobjects of X is closed under binary suprema.

Lemma 1.17 *Let $\{C_i\}_{i \in I} \subseteq CL(\mathcal{X}, \mathcal{M})$. We have that:*

(a) $\wedge C_i$ is idempotent (hereditary) if each C_i is;

(b) $\vee C_i$ is weakly hereditary (additive) if each C_i is.

This allows us to give the following:

Definition 1.18 Given a closure operator C, its *idempotent hull* $\hat{C}$ and its *hereditary hull* $\tilde{C}$ are defined by

$$\hat{C} = \wedge\{D \in iCL(\mathcal{X}, \mathcal{M}) : C \leq D\} \quad \textit{and}$$
$$\tilde{C} = \wedge\{D \in hCL(\mathcal{X}, \mathcal{M}) : C \leq D\};$$

its *weakly hereditary core* $\check{C}$ and its *additive core* C^+ are defined by

$$\check{C} = \vee\{D \in wCL(\mathcal{X}, \mathcal{M}) : D \leq C\} \quad \textit{and}$$
$$C^+ = \vee\{D \in aCL(\mathcal{X}, \mathcal{M}) : D \leq C\}.$$

As a consequence, the following result is obtained:

Theorem 1.19

(a) The conglomerate $iCL(\mathcal{X},\mathcal{M})$ [$hCL(\mathcal{X},\mathcal{M})$] is reflective in $CL(\mathcal{X},\mathcal{M})$. The reflector sends C to its idempotent [hereditary] hull.

(b) The conglomerate $wCL(\mathcal{X},\mathcal{M})$ [$aCL(\mathcal{X},\mathcal{M})$] is coreflective in $CL(\mathcal{X},\mathcal{M})$. The coreflector sends C to its weakly hereditary [additive] core.

Further insight into the relationship between a closure operator and its idempotent hull or its weakly hereditary core is brought by the following commutative diagram of Galois connections that appeared in [CKS5].

$$\begin{tikzcd}
 & & CL(\mathcal{X},\mathcal{M}) \arrow[dr, "\dot{\nabla}"] & & \\
 & wCL(\mathcal{X},\mathcal{M}) \arrow[ur, "\ddot{\Delta}"] & & iCL(\mathcal{X},\mathcal{M}) \arrow[dr, "\ddot{\nabla}"] & \\
P(\mathcal{M}) \arrow[ur, "\dot{\Delta}"] \arrow[rrrr, "\nu"'] & & & & P(\mathcal{M})^{\mathrm{op}}
\end{tikzcd}$$

The above Galois connections are as follows.

(1) $\dot{\Delta}^*$ associates to each weakly hereditary closure operator its class of C-dense $\mathcal{M}$-subobjects and $\dot{\Delta}_*$ is its corresponding coadjoint. We have that $\dot{\Delta}_* \circ \dot{\Delta}^* \simeq id_{wCL(\mathcal{X},\mathcal{M})}$.

(2) $\ddot{\Delta}^*$ associates to each closure operator its weakly hereditary core and $\ddot{\Delta}_*$ is the inclusion.

(3) $\dot{\nabla}_*$ associates to each closure operator its idempotent hull and $\dot{\nabla}^*$ is the inclusion.

(4) $\ddot{\nabla}_*$ associates to each idempotent closure operator its class of C-closed $\mathcal{M}$-subobjects and $\ddot{\nabla}^*$ is its corresponding adjoint. We have that $\ddot{\nabla}^* \circ \ddot{\nabla}_* \simeq id_{iCL(\mathcal{X},\mathcal{M})}$.

(5) ν is the polarity induced by the relation $\perp \subseteq \mathcal{M} \times \mathcal{M}$ defined by: $m \perp n$ iff for every pair of morphisms f, g such that $f \circ m = n \circ g$, there exists a unique morphism d such that both triangles of the following diagram

$$\begin{tikzcd} \bullet \arrow[r, "g"] \arrow[d, "m"'] & \bullet \arrow[d, "n"] \\ \bullet \arrow[r, "f"'] \arrow[ur, "d"] & \bullet \end{tikzcd}$$

commute.

A more complete version of the above diagram together with the relative proofs can be found in [CKS3].

As a direct consequence of the general theory of Galois connections and the above statements we obtain the following proposition.

Proposition 1.20

(1) Let ∇ be the composite $CL(\mathcal{X},\mathcal{M}) \xrightarrow{\dot{\nabla}} iCL(\mathcal{X},\mathcal{M}) \xrightarrow{\ddot{\nabla}} P(\mathcal{M})^{\mathbf{op}}$. Then, $CL(\mathcal{X},\mathcal{M}) \xrightarrow{\nabla_*} P(\mathcal{M})^{\mathbf{op}}$ and $P(\mathcal{M})^{\mathbf{op}} \xrightarrow{\nabla^*} CL(\mathcal{X},\mathcal{M})$ are given by:
$\nabla_*(C) = \{m \in \mathcal{M} : m \text{ is } C - closed\}$
$\nabla^*(\mathcal{N}) = sup\{C \in CL(\mathcal{X},\mathcal{M}) : \nabla_*(C) \supseteq \mathcal{N}\}$.

(2) Let Δ be the composite $P(\mathcal{M}) \xrightarrow{\dot{\Delta}} wCL(\mathcal{X},\mathcal{M}) \xrightarrow{\ddot{\Delta}} CL(\mathcal{X},\mathcal{M})$. Then, $P(\mathcal{M}) \xrightarrow{\Delta_*} CL(\mathcal{X},\mathcal{M})$ and $CL(\mathcal{X},\mathcal{M}) \xrightarrow{\Delta^*} P(\mathcal{M})$ are given by:
$\Delta^*(C) = \{m \in \mathcal{M} : m \text{ is } C - dense\}$
$\Delta_*(\mathcal{N}) = inf\{C \in CL(\mathcal{X},\mathcal{M}) : \Delta^*(C) \supseteq \mathcal{N}\}$.

Further expansions of the above diagram, that for space limitations are not included here, can be found in [CKS5].

A very special closure operator construction, normally referred to as the Salbany construction, naturally belongs to this section. However, due to its tight relationship with epimorphisms, we delay its introduction until next section.

2 Applications to Epimorphisms

A special case of an idempotent closure operator arises in the following way:

Definition 2.1 Given any class $\mathcal{A}$ of $\mathcal{X}$-objects and $M \xrightarrow{m} X$ in $\mathcal{M}$, define $m^{\mathcal{A}}$ to be the intersection of all equalizers of pairs of $\mathcal{X}$-morphisms r, s from X to some $\mathcal{A}$-object A that satisfy $r \circ m = s \circ m$, and let $m_{\mathcal{A}} \in \mathcal{M}$ be the unique $\mathcal{X}$-morphism by which m factors through $m^{\mathcal{A}}$.

It is easy to see that this gives rise to an idempotent closure operator that will be denoted by $S_{\mathcal{A}}$ and that will be called the *regular closure operator induced by* $\mathcal{A}$. This generalizes the Salbany construction of closure operators induced by classes of topological spaces (cf. [S]).

Properties of regular closure operators and related results can be found in [C_1], [CKS_1], [CKS_{4-5}], [CS], [DG_3], [DGT].

We illustrate this new concept with a few examples.

Example 2.2 Let $M \xrightarrow{m} X$ be an $\mathcal{M}$-subobject. In the category **Top**:

(a). If $\mathcal{A} = \mathbf{Top_0}$, then $S_{\mathbf{Top_0}}$ is the b-closure (cf. Example 1.6 (d)).

(b). If $\mathcal{A} = \mathbf{Top_1}$, then $S_{\mathbf{Top_1}}$ is the discrete closure inside $\mathbf{Top_1}$.

(c). If $\mathcal{A} = \mathbf{Top_2}$, then $S_{\mathbf{Top_2}}$ is the usual topological closure inside $\mathbf{Top_2}$.

In the category **Grp**:

(d). If $\mathcal{A} = \mathbf{Ab}$, then $m^{\mathbf{Ab}}$ is the intersection of all normal subgroups K containing M such that X/K is abelian.

In the category **Ab**:

(e). If $\mathcal{A}$ is the subcategory **Tf** of all torsion free abelian groups, then $M^{\mathbf{Tf}}$ is the intersection of all subgroups K of X containing M such that X/K is torsion free. We have that $M^{\mathbf{Tf}}$ always contains $M + tor(X)$, where $tor(X)$ denotes the torsion subgroup of X. However, equality does not usually hold.

(f). If $\mathcal{A}$ is the subcategory **Red** of all Reduced abelian groups, then $M^{\mathbf{Red}}$ is the intersection of all subgroups K of X containing M such that X/K is reduced. We have that $M^{\mathbf{Red}}$ always contains $M + div(X)$, where $div(X)$ denotes the largest divisible subgroup of X, but as in (e) equality does not usually hold.

Notice that although always idempotent, regular closure operators need not be weakly hereditary and 2.2 (d) is an example of a non-weakly hereditary one. Conditions under which a regular closure operator is weakly hereditary in some concrete categories can be found in [Cl_2] and [F_3].

Giuli was the first one to prove in [G_1] that the epimorphisms in an epireflective subcategory $\mathcal{A}$ of **Top** can be characterized by means of the corresponding regular closure operator. An intermediate step in the attempt of generalizing this result to an arbitrary category appeared in [C_1], where the setting of a concrete category over an arbitrary category $\mathcal{X}$ was used and the assumption of $\mathcal{A}$ being epireflective was removed. The final version of this result with respect to the current notion of closure operator appeared in a remark in [DG_3]. We state the complete theorem.

Theorem 2.3 *Let $\mathcal{X}$ have equalizers, let $\mathcal{M}$ contain all regular monomorphisms and let $\mathcal{A}$ be a subcategory of $\mathcal{X}$. Then, an $\mathcal{A}$-morphism $X \xrightarrow{f} Y$ is an epimorphism in $\mathcal{A}$ if and only if f is $S_{\mathcal{A}}$-dense.*

By applying this result to the above examples (a), (b) and (c) we obtain that the epimorphisms in $\mathbf{Top}_0$ are the b-dense continuous functions, in

$\mathbf{Top}_1$ they are surjective and in $\mathbf{Top}_2$ they are the usual dense continuous functions. (d) implies that the epimorphisms in **Ab** are surjective homomorphisms. The same conclusion can be drawn in the subcategory **Tor** of abelian torsion groups and in the subcategory **Div** of divisible abelian groups using (e) and (f). Moreover, if $\mathbf{2Z} \xrightarrow{\mathbf{i}} \mathbf{Z}$ is the inclusion of the even integers into the integers, then from (e) we conclude that i is $S_{\mathbf{Tf}}$-dense and so an epimorphisms in **Tf**. Thus in this category the epimorphisms are not surjective. As a matter of fact, a **Tf**-morphism $X \xrightarrow{f} Y$ is an epimorphism in **Tf** if and only if $Y/f(X)$ is a torsion group. A similar conclusion can be drawn about the subcategory **Red** where we have that a **Red**-morphism $X \xrightarrow{f} Y$ is an epimorphism in **Red** if and only if $Y/f(X)$ is a divisible group. For further examples cf. [C_1]. Moreover, a dual notion of closure operator (termed co-closure operator) was also introduced in [C_1] and used to obtain a characterization of the monomorphisms in a subcategory $\mathcal{A}$ in terms of the co-closure.

We would like to observe that the regular closure operator induced by a subcategory $\mathcal{A}$ was also used in **Top** to analyze the co-well poweredness of $\mathcal{A}$ or some other related category. However we will not digress on this point and we refer the reader for more details to [DG_2], [GH_1] and [DGT].

3 Applications to Separation

We recall that a topological space X is called a Hausdorff space if its diagonal Δ_X is closed in $X \times X$. Dikranjan and Giuli were the first ones to notice in [DG_1] that if the word "closed" would be replaced by "C-closed" where C is an appropriate regular closure operator, then the above result could be extended to several subcategories of topological spaces. The following general version of this result for epireflective subcategories of **Top** was obtained in [GH_1].

Theorem 3.1 *Let $\mathcal{A}$ be an epireflective subcategory of* **Top** *and let $Q(\mathcal{A})$ be its quotient reflective hull. A topological space X belongs to $Q(\mathcal{A})$ iff the diagonal Δ_X is $S_{\mathcal{A}}$-closed in $X \times X$.*

In [CS], a first attempt to define a notion of closure operator in a more general setting was made. The setting of a concrete functor $\mathcal{A} \xrightarrow{U} \mathcal{X}$ was used together with a class of $\mathcal{X}$-monomorphisms $\mathcal{M}$ with the assumption that pullbacks and arbitrary intersections of regular subobjects exist in $\mathcal{X}$ and $\mathcal{M}$ is closed under these constructions. Moreover, $\mathcal{X}$ is assumed to have squares and equalizers. The practical difference between the notion of *global closure operator* introduced in this paper and the current notion of closure operator is that idempotency was already included in the definition and the morphism-consistency condition 1.2 (c) was given as follows: for every morphism $X \xrightarrow{f} Y$, $f^{-1}(m)$ is C_X-closed whenever m is C_Y-closed. This is in general weaker than 1.2 (c). However for idempotent closure operators, the two conditions coincide.

Sep(C) will denote all those objects $Y \in \mathcal{A}$ with the property that for every $X \in \mathcal{A}$, $m \in \mathcal{M}_{UX}$ and morphisms $f, g : X \longrightarrow Y$ such that $Uf \circ m = Ug \circ m$, we have that $\overline{Uf \circ m}^{C} = \overline{Ug \circ m}^{C}$.

For a subcategory $\mathcal{B}$ of $\mathcal{A}$, the *closure operator hull of* $\mathcal{B}$ was defined as $Sep(S_{\mathcal{B}})$.

The following definition was given in [CS].

Definition 3.2 For every $X \in \mathcal{A}$, $(\Delta_X, \delta_X) = equ(U\pi_1, U\pi_2)$, with π_1 and π_2 being the usual projections from $X \times X$ into X, is called the *diagonal of* $X \times X$.

As a consequence of the above setup, the following two results were obtained.

Theorem 3.3 *Let $\mathcal{A}$ have squares and let U preserve them. Then for every global closure operator C we have that $Sep(C) = \{Y \in \mathcal{A}$ such that Δ_Y is C-closed in $Y \times Y\}$.*

Corollary 3.4 *Let $\mathcal{A}$ have squares and let U preserve them. Then for every $Y \in \mathcal{A}$ and for every subcategory $\mathcal{B}$ of $\mathcal{A}$, Y belongs to the closure operator hull of $\mathcal{B}$ iff Δ_Y is $S_{\mathcal{B}}$-closed in $Y \times Y$.*

It is worth mentioning that in this paper the regular closure operator $S_{\mathcal{B}}$ for $\mathcal{B} \subseteq \mathcal{A}$ was used to produce a Galois connection between global closure operators on $\mathcal{A}$ and subcategories of $\mathcal{A}$. Subsequently this was used to provide a factorization of the Pumplün–Röhrl connection (cf. [CKS$_1$], [PR]).

After the current version of categorical closure operator was introduced in [DG$_3$] a very general version of the above corollary that by now was called the "diagonal theorem" appeared in [GMT]. In order to present this result we need to introduce a few concepts.

Throughout the remainder of this section we will no longer consider a concrete functor $\mathcal{A} \xrightarrow{U} \mathcal{X}$ but simply an arbitrary category $\mathcal{X}$ with finite products and equalizers.

Definition 3.5 For a subcategory $\mathcal{A}$ of $\mathcal{X}$, a morphism $M \xrightarrow{m} X$ is called $\mathcal{A}$-regular if it is the equalizer of a pair of morphisms $f, g : X \longrightarrow A$, with $A \in \mathcal{A}$.

We also recall here the standard definition of diagonal in the category $\mathcal{X}$.

Definition 3.6 For every $X \in \mathcal{X}$, $(\Delta_X, \delta_X) = equ(\pi_1, \pi_2)$, with π_1 and π_2 being the usual projections from $X \times X$ into X, is called the *diagonal of* X.

For $\mathcal{A} \subseteq \mathcal{X}$, consider the subcategory $S(\mathcal{A}) = \{X \in \mathcal{X} :$ there is a monomorphism $X \xrightarrow{p} A$ with $A \in \mathcal{A}\}$.

Notice that if $\mathcal{A}$ is reflective in $\mathcal{X}$, then $S(\mathcal{A})$ can be described as follows: $S(\mathcal{A}) = \{X \in \mathcal{X} :$ the reflection r_X is a monomorphism$\}$.

It may also be worth observing that if $\mathcal{X}$ has (strong epi,mono)-factorizations then $S(\mathcal{A})$ is the *strongly epireflective hull of* $\mathcal{A}$.

For Y, X in $\mathcal{X}$, let $k_{Y,X} : r(Y \times X) \longrightarrow rY \times rX$ be the canonical morphism induced by the reflection and the product construction. Then we have the following:

Theorem 3.7 *For a reflective subcategory $\mathcal{A}$ of $\mathcal{X}$ and an object $X \in \mathcal{X}$, assume that the canonical morphism $k_{Y,X}$ is a monomorphism for every $Y \in \mathcal{X}$. Then, X belongs to $S(\mathcal{A})$ if and only if Δ_X is $\mathcal{A}$-regular.*

For additional results concerning the concept of diagonal we refer to [GMT] and [T_1].

Further development of a general notion of separation appeared in [CT_1]. The setup in that paper is the one of a finitely complete $(\mathcal{E}, \mathcal{M})$-category $\mathcal{X}$ in which $\mathcal{E}$ is a class of epimorphisms and $\mathcal{M}$ is a class of monomorphisms, both containing the $\mathcal{X}$-isomorphisms. It is also assumed that $\mathcal{X}$ is $\mathcal{M}$-complete so that $\mathcal{X}$ has multiple pullbacks of arbitrary sinks of $\mathcal{M}$-morphisms. It is well known that in this case, the above factorization structure for morphisms $(\mathcal{E}, \mathcal{M})$ extends to a factorization structure for sinks $(\mathbf{E}, \mathcal{M})$. We observe that since $\mathcal{E}$ is a class of epimorphisms, $\mathcal{M}$ must contain all regular monomorphisms.

In order to present an interesting result from this paper, we need to introduce the following concepts.

Definition 3.8

(a) A $\mathcal{X}$-object X is *preterminal* if for all $\mathcal{X}$-morphisms $h, k : Z \longrightarrow X$ one has $h = k$.

(b) A *prepoint* of an object X is an $\mathcal{M}$-subobject $P \xrightarrow{p} X$ with P preterminal; it is a *quasipoint* (*point*) of X if $P \simeq X_t$ ($P \simeq T$), where X_t is the middle object of the $(\mathcal{E}, \mathcal{M})$-factorization of the morphism $X \xrightarrow{t} T$ with T being a terminal object.

(c) *Quasipoints detect monosources* if an $\mathcal{X}$-source $(X \xrightarrow{f_i} Y_i)_{i \in I}$ is monic whenever the source $\mathcal{X}(X_t, X) \xrightarrow{\mathcal{X}(X_t, f_i)} \mathcal{X}(X_t, Y_i)_{i \in I}$ is monic in $\mathbf{Set}$.

Notice that since $\mathcal{M}$ contains all regular monomorphisms, $(\Delta_X, \delta_X) \in \mathcal{M}$ for each $X \in \mathcal{X}$. So, the following definition can be given:

Definition 3.9 Let C be a closure operator on $\mathcal{X}$. A $\mathcal{X}$-object X is called C-separated if its diagonal (Δ_X, δ_X) is C-closed.

For a closure operator C, $\Delta(C)$ will denote the full subcategory of $\mathcal{X}$ whose objects are C-separated.

We conclude this section with the following:

Theorem 3.10 *Let quasipoints detect monosources of $\mathcal{X}$. Then a full subcategory $\mathcal{A}$ of $\mathcal{X}$ is of the form $\Delta(C)$ for some C if and only if $\mathcal{A}$ is closed under monosources in $\mathcal{X}$.*

4 Applications to Compactness

Herrlich, Salicrup and Strecker in [HSS] presented a generalization of the classical notion of compactness for topological spaces. Such a generalization provides in an abstract category $\mathcal{X}$ a concept of compactness with respect to a factorization structure on $\mathcal{X}$. This notion was successfully used by Fay in [F_1].

As a consequence of the bijective correspondence between factorization structures on $\mathcal{X}$ and weakly hereditary and idempotent closure operators on $\mathcal{X}$ (cf. Theorem 1.14) one naturally obtains a notion of compactness with respect to such closure operators. In order to extend this notion to those closure operators that do not satisfy the weakly hereditary and idempotent conditions, a notion of compactness directly with respect to a closure operator was introduced nearly at the same time in [DG_5] and [C_3]. However, the abstract category setup in [C_3] was much more general than the category of convergence spaces used in [DG_5]. This last paper was followed by [DG_6], but only in [CGT_1] were the ideas introduced in [DG_5] generalized to an appropriate abstract setting that encompasses all previous attempts, including [C_3]. Therefore here we introduce the topic with the results in [C_3] and then conclude it with the ones in [CGT_1], followed by a number of examples.

Here the setup is exactly as in Section 1 and C will denote a closure operator as in Definition 1.2.

Definition 4.1 A $\mathcal{X}$-morphism $X \xrightarrow{f} Y$ is called *C-closed preserving* if for every C-closed $\mathcal{M}$-subobject $M \xrightarrow{m} X$, in the $(\mathbf{E}, \mathcal{M})$-factorization $m_1 \circ e_1 = f \circ m$, m_1 is C-closed.

Definition 4.2 A $\mathcal{X}$-object X is called C-compact if for each $\mathcal{X}$-object Z, the projection $X \times Z \xrightarrow{p_Z} Z$ is C-closed preserving.

$Comp(C)$ will denote the subcategory of $\mathcal{X}$ whose objects are the C-compact ones.

Clearly, this definition generalizes the classical notion of compactness in topology since if $\mathcal{X}$=**Top** and C is the Kuratowski closure induced by the topology, then $Comp(C)$ consists exactly of the compact topological spaces.

We include here a series of results that clearly generalize classical properties of topological compactness.

Theorem 4.3

(a) *Comp(C) is closed under the formation of finite products.*

(b) *Let $\mathcal{X}$ have products and suppose that the pullback of $\Pi X_i \times Z \xrightarrow{\pi_i \times 1_Z} X_i \times Z$ along any C-closed subobject belongs to $\mathbf{E}$ for every $i \in I$. Then, if the ΠX_i is C-compact, so is each factor X_i.*

(c) *If C is weakly hereditary and idempotent, then the C-compact objects are closed under the formation of C-closed $\mathcal{M}$-subobjects.*

(d) *Suppose that for $e \in \mathbf{E}$, the pullback of $e \times 1$ along any C-closed subobject belongs to $\mathbf{E}$. If $X \xrightarrow{f} Y$ is a $\mathcal{X}$-morphism and (e, m) is its $(\mathbf{E}, \mathcal{M})$-factorization, then if X is C-compact, so is X_f (where X_f denotes the middle object of the $(\mathbf{E}, \mathcal{M})$-factorization).*

Definition 4.4 Let $X \xrightarrow{f} Y$ be a $\mathcal{X}$-morphism. The morphism $X \xrightarrow{<1_X, f>} X \times Y$ is called the *graph of f*.

It is well known that the morphism $X \xrightarrow{<1_X, f>} X \times Y$ can be seen as the pullback of (Δ_Y, δ_Y) along the morphism $f \times 1_Y$. So, if $\mathcal{M}$ contains all regular monomorphisms, then $< 1_X, f > \in \mathcal{M}$ for every $X \in \mathcal{X}$.

Proposition 4.5 *Let $\mathcal{X}$ have equalizers and let $\mathcal{M}$ contain all regular monomorphisms. A $\mathcal{X}$-object Y is C-separated iff for every morphism $X \xrightarrow{f} Y$, $(X, < 1_X, f >)$ is C-closed in $X \times Y$.*

The previous result, besides having its own merits, was also useful in proving the following:

Theorem 4.6 *Let $\mathcal{X}$ have equalizers and let $\mathcal{M}$ contain all regular monomorphisms. A C-compact subobject of a C-separated object is C-closed.*

Corollary 4.7 *Let $\mathcal{X}$ have equalizers and let $\mathcal{M}$ contain all regular monomorphisms. Assume that C is weakly hereditary and idempotent and that for $e \in \mathbf{E}$, the pullback of $e \times 1$ along any C-closed subobject belongs to $\mathbf{E}$. If X is C-compact and Y is C-separated, then any $\mathcal{X}$-morphism $X \xrightarrow{f} Y$ is C-closed preserving.*

An attempt towards a generalization of the classical Čech–Stone compactification also appeared in [C3]. We briefly recall it here.

Let us assume that $\mathcal{X}$ has equalizers and that $\mathcal{M}$ contains all regular monomorphisms. As a consequence we have that $\mathbf{E}$ is a class of episinks. We also assume that $\mathcal{X}$ is an $\mathbf{E}$-co-well powered category with products.

Definition 4.8 A *C-compactification* of an object $X \in \Delta(C)$ is a pair $(\beta, \beta_C X) \in Comp(C) \cap \Delta(C)$ with $X \xrightarrow{\beta} \beta_C X$ $S_{\Delta(C)}$-dense and such that for every morphism $X \xrightarrow{f} Y$ with $Y \in Comp(C) \cap \Delta(C)$, there exists a morphism $\beta_C X \xrightarrow{g} Y$ with $g \circ \beta = f$.

We conclude the list of results from [C$_3$] with the following:

Theorem 4.9 *Let C be an idempotent and weakly hereditary closure operator such that $\Delta(C)$ is co-well powered. If $Comp(C)$ is closed under the formation of products, then every $X \in \Delta(C)$ has a C-compactification.*

Conditions under which $Comp(C)$ is closed under products can be found in [CGT$_1$] and [CT$_2$] and will be presented later in this paper.

Further results about compactness with respect to regular closure operators in some concrete categories can be found in [C$_4$].

As already mentioned above, the theory of compactness with respect to a closure operator reached its present (and probably final) stage in [CGT$_1$]. In that paper, the setup of a complete category $\mathcal{X}$ with a proper $(\mathcal{E}, \mathcal{M})$-factorization structure for morphisms is used. That is, besides the usual properties, $\mathcal{E}$ is assumed to be a class of epimorphisms and $\mathcal{M}$ a class of monomorphisms. $\mathcal{X}$ is also assumed to have multiple pullbacks of arbitrary large families of morphisms in $\mathcal{M}$ with common codomain, with the pullbacks belonging to $\mathcal{M}$. This is known to be equivalent to the fact that the $(\mathcal{E}, \mathcal{M})$-factorization structure for morphisms extends to an $(\mathbf{E}, \mathcal{M})$-factorization structure for sinks (cf. [AHS]). Some consequences of this have already been mentioned in Proposition 1.1.

As usual, C will denote a closure operator on $\mathcal{X}$ with respect to $\mathcal{M}$.

The approach to a general notion of compactness in [CGT$_1$] begins with the following definitions.

Definition 4.10 A $\mathcal{X}$-morphism $X \xrightarrow{f} Y$ is called C-preserving if for every $\mathcal{M}$-subobject $M \xrightarrow{m} X$, $(m^C)_f \simeq (m_f)^C$.

Definition 4.11 A $\mathcal{X}$-object X is called C-compact if the projection $X \times Y \xrightarrow{p_Y} Y$ is C-preserving for every $\mathcal{X}$-object Y.

Notice that due to the fact that any C-preserving morphism sends C-closed subobjects into C-closed subobjects, the above notion of C-compactness is stronger than the one given in Definition 4.2. However, if C is idempotent, then the two notions are equivalent.

Here we will keep using the symbol $Comp(C)$ to denote all C-compact objects with respect to the above definition.

We refrain from including properties of C-preserving morphisms for which we refer the reader to [CGT$_1$] and we go directly to the results about C-compactness.

Proposition 4.12

(a) If $X \in \mathcal{X}$ is C-compact and $M \xrightarrow{m} X$ in $\mathcal{M}$ is C-closed, with C weakly hereditary, then M is C-compact.

(b) For X C-compact and Y C-separated, every morphism $X \xrightarrow{f} Y$ is C-preserving.

(c) For $X \xrightarrow{f} Y$ in $\mathcal{E}$, with $\mathcal{E}$ stable under pullbacks, if X is C-compact, so is Y.

(d) $Comp(C)$ is closed under finite products in $\mathcal{X}$.

A very interesting novelty about this paper is that one can develop a theory of compactness for morphisms entirely for free, by simply reinterpreting the results already obtained for objects in the comma categories of the given category. In what follows we outline this idea.

Clearly any closure operator C on $\mathcal{X}$ with respect to $\mathcal{M}$ induces a closure operator C^Y on the comma category $\mathcal{X}/Y$ with respect to $\mathcal{M}_Y$ as follows: for $Y \in \mathcal{X}$ and $h \xrightarrow{m} f$ in $\mathcal{M}_Y$ with $X \xrightarrow{f} Y$ let $C^Y_f(m) = C_X(m) : f \circ C_X(m) \longrightarrow f$, i.e., closures are formed as in $\mathcal{X}$. It follows easily that if C is idempotent or weakly hereditary so is C^Y.

Now we can give the following definition.

Definition 4.13

(a) A $\mathcal{X}$-morphism $A \xrightarrow{f} B$ is C-compact if f is C^B-compact as an object of $\mathcal{X}/B$.

(b) A $\mathcal{X}$-morphism $A \xrightarrow{f} B$ is C-separated if f is C^B-separated as an object of $\mathcal{X}/B$.

(c) A $\mathcal{X}$-morphism $A \xrightarrow{f} B$ is C-perfect if it is C-compact and C-separated.

We observe that (a) is equivalent to: in every pullback diagram

$$\begin{tikzcd} X \arrow[r, "f'"] \arrow[d, "h'"'] & Y \arrow[d, "h"] \\ A \arrow[r, "f"'] & B \end{tikzcd}$$

in $\mathcal{X}$, f' is C-preserving. Moreover (b) is equivalent to: the diagonal morphism $A \xrightarrow{\delta_f} A \times_B A$ is C-closed. Please, notice that in [CGT_1], the terminology C-Hausdorff was used instead of C-separated.

The following results were proved.

Proposition 4.14 *If in the above pullback diagram Y and f are C-compact, then so is X. In particular, a C-compact morphism is C-preserving and has C-compact fibres.*

Proposition 4.15

(a) Let C be hereditary and let $\mathcal{E}$ be stable under pullbacks along $\mathcal{M}$-morphisms. Then, a morphism $X \xrightarrow{f} Y$ is C-compact if and only if $X \times Z \xrightarrow{f \times 1_Z} Y \times Z$ is C-preserving for every object Z.

(b) Under the hypotheses of part (a), every morphism $X \xrightarrow{f} Y$ with X C-compact and Y C-separated is C-compact.

Proposition 4.16

(a) Every isomorphism is C-compact and every C-closed morphism in $\mathcal{M}$ is C-compact if C is weakly hereditary.

(b) The class of C-compact morphisms in $\mathcal{X}$ is stable under pullbacks, closed under composition and closed under the formation of finite direct products.

(c) Let the composite $g \circ f$ be C-compact. Then:

(i) f is C-preserving if g is C-separated;

(ii) f is C-compact if g is a monomorphism;

(iii) g is C-compact if $f \in \mathcal{E}$, with $\mathcal{E}$ stable under pullbacks.

The basic properties of C-perfect morphisms can be summarized as follows.

Proposition 4.17 *The class of C-perfect morphisms in $\mathcal{X}$ contains all isomorphisms, even all C-closed morphisms of $\mathcal{M}$ if C is weakly hereditary. It is stable under pullbacks, left cancellable with respect to monomorphism and closed under the formation of finite direct products. It is also closed under composition if C is weakly hereditary.*

The following definitions will be needed in order to present the final results.

Definition 4.18 We say that the class of morphisms $\mathcal{E}$ is a *surjectivity class* if there exists a class $\mathcal{P}$ of $\mathcal{X}$-objects such that a $\mathcal{X}$-morphism $X \xrightarrow{f} Y$ belongs to $\mathcal{E}$ if and only if every $P \in \mathcal{P}$ is projective with respect to f (that is every morphism $P \xrightarrow{g} Y$ factors as $g = f \circ h$ for some morphism $P \xrightarrow{h} X$.)

Definition 4.19 For a small family $(X_i)_{i\in I}$ of $\mathcal{X}$-objects and any subset $J \subseteq I$, let $X_J = \prod_{i\in J} X_i$ and let $X_I \xrightarrow{\pi_J} X_J$ denote the corresponding projection. A closure operator C is said to satisfy the *finite structure property of products* (FSPP) if for all $\mathcal{M}$-subobjects m, n of X_I, one has that if $\pi_F(n) \leq C_{X_F}(\pi_F(m))$ holds for every finite subset $F \subseteq I$, then $n \leq C_{X_I}(m)$.

It is easy to see that this is equivalent to: for every $\mathcal{M}$-subobject $M \xrightarrow{m} X_I$, $C_{X_I}(m) \simeq \bigcap_F \pi_F^{-1}(C_{X_F}(\pi_F(m)))$, where F runs over the finite subsets of I.

Remark 4.20 It is well known that the product X_I can be seen as an inverse limit of the finite products X_F (cf. [HS, Proposition 22.5]). As observed in [CGT$_1$, 2.7] the purpose of FSPP is to obtain that for every $\mathcal{M}$-subobject $M \xrightarrow{m} X_I$, $C_{X_I}(m)$ can be seen as an inverse limit of the family $(C_{X_F}(\pi_F(m)))_{F\subseteq I}$ with F running over the finite subsets of I.

The following generalization of two classical results is obtained.

Theorem 4.21 (Tychonoff's Theorem) *If $\mathcal{E}$ is a surjectivity class and C has FSPP, under the Axiom of Choice, Comp(C) is closed under direct products in $\mathcal{X}$.*

Theorem 4.22 (Frolík's Theorem) *Let $\mathcal{E}$ be a surjectivity class and let C be a hereditary closure operator with FSPP. Then the direct product $\Pi f_i : \Pi X_i \longrightarrow \Pi Y_i$ of a family $(X_i \xrightarrow{f_i} Y_i)_{i\in I}$ of C-compact (C-perfect) morphisms is C-compact (C-perfect).*

A full subcategory $\mathcal{A}$ of $\mathcal{X}$ is called *C-co-well powered* if every $A \in \mathcal{A}$ has only a small set of non-isomorphic C-dense morphisms with domain A and codomain in $\mathcal{A}$.

We conclude this section with the following result that provides an extension of the Čech–Stone compactification.

Theorem 4.23 *Assume that C is a weakly hereditary and idempotent closure operator with FSPP and that $\mathcal{E}$ is a surjectivity class. Then, if $\Delta(C)$ is C-co-well powered, then Comp(C)$\cap\Delta(C)$ is C-dense reflective in $\Delta(C)$.*

Example 4.24

(a) In **Top**, if C is the Kuratowski closure, then *Comp*(C) consists of the classical compact topological spaces and the preceding theorem gives the Čech–Stone compactification of a Hausdorff space.

(b) Since for any bireflective subcategory $\mathcal{B}$ of **Top**, $S_{\mathcal{B}}$ is the identity, then *Comp*($S_{\mathcal{B}}$)=**Top**.

(c) *Comp*($S_{\mathbf{Top_0}}$)= b-compact topological spaces, [DG$_5$, Example 3.2].

(d) $Comp(S_{\mathbf{Top_1}})$= **Top**, (cf. Example 2.2 (b)).

(e) Consider the Θ-closure in **Top**. Then $Comp(\Theta)$ consists of all H-closed spaces [DG_6, Example 4.7(f)].

(f) Let Σ denote the sequential closure in **Top**. Then $Comp(\Sigma)$ is the class of all sequentially compact topological spaces [DG_6, Example 4.7(g)].

(g) In **Ab** let **Tf** be the subcategory of torsion-free abelian groups. Then $Comp(S_{\mathbf{Tf}}) \cap \Delta(S_{\mathbf{Tf}})$ is the subcategory **Tfd** of torsion-free divisible abelian groups. Moreover if **Z** is for instance the additive group of integers, then its $S_{\mathbf{Tf}}$-compactification is the additive group of rationals **Q** [C_3, Example 3.7].

(h) In **Ab** let $\mathcal{F} = \{G : G^1 = 0\}$ where G^1 denotes the first Ulm subgroup of G and let $\mathcal{F}_p = \{G : p^\omega G = 0\}$. Then we have that an abelian group G is $S_{\mathcal{F}}$-compact ($S_{\mathcal{F}_p}$-compact) iff G/G^1 ($G/p^\omega G$) is compact in its **Z**-adic (p-adic) topology, [FS, Theorem 3].

(i) Let **ModR** be the category of left **R**-modules over a ring **R** with unity and let r be a pre-radical, i.e., a subfunctor of the identity functor $1_{\mathbf{ModR}}$. For every $X \in \mathbf{ModR}$ and $M \xrightarrow{m} X$ define the following two closure operators: $(C_r)_X(M) = M + r(X)$ and $(C^r)_X(M) = q^{-1}(r(X/M))$, where $X \xrightarrow{q} X/M$ is the quotient map. Then we have that $Comp(C_r) = \mathbf{ModR}$ and $Comp(C^r)$ consists of all left **R**-modules X such that for each C^r-closed submodule M of X, X/M is absolutely C^r-closed, [DG_6, Example 4.7(k)]. In particular, if **R**= **Z** and r is the torsion pre-radical then we have that M is C^r-compact iff $M/r(M)$ is divisible [F_1, Example 2.4].

The reader interested in further examples could check [DG_{4-6}], [F_{1-2}], [FJ], [FS], [FW_{1-6}].

5 Applications to Connectedness

The development of a general theory of topological connectedness was started by Preuß (cf. [Pr_{1-3}]) and by Herrlich [H]. Afterwards, a considerable number of papers have been published on this subject and on possible generalizations of it (e.g., [AW], [HP], [P], [SV] and [T_2]). Most of these papers used the common approach of first defining a notion of constant morphism and then using it to introduce the notions of connectedness and disconnectedness, accordingly. Castellini and Hajek in [CH] were the first ones to introduce a notion of connectedness with respect to a closure operator. The setup of this paper was a concrete category over **Set**.

Subsequently, the theory was taken a step forward by the introduction of a similar notion in an arbitrary category **X**. This more general notion was studied in [C_{5-8}] and it is currently under investigation (cf. [C_{9-10}]).

Independently, a notion of connectedness with respect to a closure operator was introduced in [CT_1] and further developed in [Cl_5].

Since a unified theory on this subject does not exist yet, we will present the main results of the two approaches.

Since the theory presented in [CH] is superseded by the more general approach that appears in [C_{5-8}], we will introduce this last one directly.

Here we consider a category $\mathcal{X}$ together with a fixed class $\mathcal{M}$ of $\mathcal{X}$-monomorphisms and a class $\mathbf{E}$ of $\mathcal{X}$-sinks such that $\mathcal{X}$ is an $(\mathbf{E}, \mathcal{M})$-*category for sinks.*

Let $S(\mathcal{X})$ be the collection of all subcategories of $\mathcal{X}$, ordered by inclusion and let $\mathcal{N}$ be a fixed subclass of $\mathcal{M}$. For every $X \in \mathcal{X}$, we denote by $\mathcal{N}_X$ all the $\mathcal{N}$-subobjects that have X as codomain.

We begin with the following:

Definition 5.1 ([C_5]) A $\mathcal{X}$-morphism $X \xrightarrow{f} Y$ is called $\mathcal{N}$-constant if for every $\mathcal{N}$-subobject $N \xrightarrow{n} X$, we have that $n_f \simeq (id_X)_f$.

This immediately induces the following:

Proposition 5.2 (cf. [H]) *Define* $S(\mathcal{X}) \xrightarrow{\Delta_{\mathcal{N}}} S(\mathcal{X})^{\mathrm{op}}$ *and* $S(\mathcal{X})^{\mathrm{op}} \xrightarrow{\nabla_{\mathcal{N}}} S(\mathcal{X})$ *as follows:*

$\nabla_{\mathcal{N}}(\mathcal{A}) = \{X \in \mathcal{X} : \forall Y \in \mathcal{A}, X \xrightarrow{f} Y \text{ is } \mathcal{N}\text{-constant}\}$,

$\Delta_{\mathcal{N}}(\mathcal{B}) = \{Y \in \mathcal{X} : \forall X \in \mathcal{B}, X \xrightarrow{f} Y \text{ is } \mathcal{N}\text{-constant}\}$.

Then, $S(\mathcal{X}) \underset{\nabla_{\mathcal{N}}}{\overset{\Delta_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ *is a Galois connection.*

The following two Galois connections were introduced in [C_5].

Proposition 5.3 *Let* $CL(\mathcal{X}, \mathcal{M}) \xrightarrow{D_{\mathcal{N}}} S(\mathcal{X})^{\mathrm{op}}$ *and* $S(\mathcal{X})^{\mathrm{op}} \xrightarrow{T_{\mathcal{N}}} CL(\mathcal{X}, \mathcal{M})$ *be defined by:*

$$D_{\mathcal{N}}(C) = \{X \in \mathcal{X} : \text{ every } n \in \mathcal{N}_X \text{ is } C\text{-closed}\},$$

$$T_{\mathcal{N}}(\mathcal{A}) = Sup\{C \in CL(\mathcal{X}, \mathcal{M}) : D_{\mathcal{N}}(C) \supseteq \mathcal{A}\}.$$

Then, $CL(\mathcal{X}, \mathcal{M}) \underset{T_{\mathcal{N}}}{\overset{D_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ *is a Galois connection.*

Proposition 5.4 *Let* $CL(\mathcal{X}, \mathcal{M}) \xrightarrow{I_{\mathcal{N}}} S(\mathcal{X})$ *and* $S(\mathcal{X}) \xrightarrow{J_{\mathcal{N}}} CL(\mathcal{X}, \mathcal{M})$ *be defined by:*

$$I_{\mathcal{N}}(C) = \{X \in \mathcal{X} : \text{ every } n \in \mathcal{N}_X \text{ is } C\text{-dense}\},$$

$$J_{\mathcal{N}}(\mathcal{B}) = Inf\{C \in CL(\mathcal{X}, \mathcal{M}) : I_{\mathcal{N}}(C) \supseteq \mathcal{B}\}.$$

stopped Then, $S(\mathcal{X}) \underset{I_{\mathcal{N}}}{\overset{J_{\mathcal{N}}}{\rightleftarrows}} CL(\mathcal{X}, \mathcal{M})$ *is a Galois connection.*

Clearly, we have the following:

Corollary 5.5 *The composition functions* $D_{\mathcal{N}} \circ J_{\mathcal{N}}$ *and* $I_{\mathcal{N}} \circ T_{\mathcal{N}}$ *give rise to a Galois connection between* $S(\mathcal{X})$ *and* $S(\mathcal{X})^{\mathbf{op}}$.

In [C_5] we presented some characterizations of the functions $T_{\mathcal{N}}$ and $J_{\mathcal{N}}$ that play a very important role throughout the development of this theory. Therefore, for reference purposes we collect them under the following:

Proposition 5.6 *For every* $\mathcal{A} \in S(\mathcal{X})^{\mathbf{op}}$ *and* $\mathcal{M}$*-subobject* $M \xrightarrow{m} X$, *with* $X \in \mathcal{X}$, *we have that*

$$m^{T_{\mathcal{N}}(\mathcal{A})} = \cap\{f^{-1}(n) : Y \in \mathcal{A}, X \xrightarrow{f} Y, N \xrightarrow{n} Y \in \mathcal{N}_Y \, and \; m \leq f^{-1}(n)\},$$

and, for every $\mathcal{B} \in S(\mathcal{X})$ *and* $\mathcal{M}$*-subobject* $M \xrightarrow{m} Y$, *with* $Y \in \mathcal{X}$, *we have that*

$$m^{J_{\mathcal{N}}(\mathcal{B})} =$$

$$sup\left(\{m\} \cup \{(id_X)_f : X \in \mathcal{B}, X \xrightarrow{f} Y \text{ and } \exists n \in \mathcal{N}_X \text{ with } n_f \leq m\}\right).$$

Moreover, $T_{\mathcal{N}}(\mathcal{A})$ *is always idempotent and* $J_{\mathcal{N}}(\mathcal{B})$ *is always weakly hereditary.*

It may be worthwhile to observe that the real purpose of Definition 5.1 was not that of generalizing the notion of constant function to an arbitrary category but instead to provide a suitable general notion of constant morphism for which the following result would hold.

Theorem 5.7 *Let* $\mathcal{N}$ *be a subclass of* $\mathcal{M}$ *closed under the formation of direct images. Then the Galois connection* $S(\mathcal{X}) \underset{\nabla_{\mathcal{N}}}{\overset{\Delta_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathbf{op}}$ *factors through* $CL(\mathcal{X}, \mathcal{M})$ *via the two Galois connections* $S(\mathcal{X}) \underset{I_{\mathcal{N}}}{\overset{J_{\mathcal{N}}}{\rightleftarrows}} CL(\mathcal{X}, \mathcal{M})$ *and* $CL(\mathcal{X}, \mathcal{M}) \underset{T_{\mathcal{N}}}{\overset{D_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathbf{op}}$.

Closure properties of the subclasses $\nabla_{\mathcal{N}}(\mathcal{A})$ and $\Delta_{\mathcal{N}}(\mathcal{B})$ for $\mathcal{A} \in S(\mathcal{X})^{\mathbf{op}}$ and $\mathcal{B} \in S(\mathcal{X})$ were studied in [C_6].

Example 5.8 (cf. [CH]) Let $\mathcal{X}$ be the category **Top** with the (episink,embedding)-factorization structure and let $\mathcal{N}$ be the class of all nonempty embeddings. Notice that since $\mathcal{N}$ contains all singleton monomorphisms (i.e., morphisms with singleton domain), to say that a morphism $X \xrightarrow{f} Y$ is $\mathcal{N}$-constant simply means that $f(X)$ is a singleton.

(a). If C is the closure operator induced by the topology, then the class $D_{\mathcal{N}}(C)$ agrees with the class **Discr** of discrete topological spaces and $\nabla_{\mathcal{N}}(\mathbf{Discr})$ consists of the classical connected topological spaces.

If $M \xrightarrow{m} X$ is an $\mathcal{M}$-subobject of $X \in \mathbf{Top}$, then $M^{T_{\mathcal{N}}(\mathbf{Discr})}$ equals the intersection of all clopen subsets of X containing M. Since $\mathcal{M}$ satisfies the conditions of Theorem 5.7, we have that the class $(I_{\mathcal{N}} \circ T_{\mathcal{N}})(\mathbf{Discr})$ consists of all connected topological spaces.

Now, let $\mathcal{B}$ be the class of all connected topological spaces. From Proposition 5.6, $M^{J_{\mathcal{N}}(\mathcal{B})}$ is the union of M with all connected subsets of X which intersect M and the subcategory of all totally disconnected topological spaces agrees with $(D_{\mathcal{N}} \circ J_{\mathcal{N}})(\mathcal{B})$. Thus, again from Theorem 5.7, connected topological spaces and totally disconnected topological spaces are fixed points of the Galois connection $(\Delta_{\mathcal{N}}, \nabla_{\mathcal{N}})$ of Proposition 5.2.

(b). Let $\mathcal{A} = \mathbf{Top_0} \in S(\mathcal{X})^{\mathrm{op}}$. **Ind** and $\mathbf{Top_0}$ are corresponding fixed points of the Galois connection $(\Delta_{\mathcal{N}}, \nabla_{\mathcal{N}})$ of Proposition 5.2 (cf. [AW]).

Let $M \xrightarrow{m} X$ be an $\mathcal{M}$-subobject of $X \in \mathbf{Top}$ and let $c(M) = \{y \in X : \exists x \in M \text{ } \mathit{with} \text{ } \overline{\{x\}} = \overline{\{y\}}\}$, where $\overline{\{x\}}$ denotes the usual topological closure of $\{x\}$. If $X \xrightarrow{r_0} r_0X$ is the $\mathbf{Top_0}$-reflection, then $M^{T_{\mathcal{N}}(\mathbf{Top_0})} = c(M) = r_0^{-1}r_0(M)$. Moreover, $M^{T_{\mathcal{N}}(\mathbf{Top_0})} \subseteq b(M)$, where $b(M)$ is the **b**-closure of M. $M^{J_{\mathcal{N}}(\mathbf{Ind})}$ is the union of M with all indiscrete subobjects of X which intersect M and $(D_{\mathcal{N}} \circ J_{\mathcal{N}})(\mathbf{Ind}) = \mathbf{Top_0}$.

(c). Let $\mathcal{A} = \mathbf{Top_1} \in S(\mathcal{X})^{\mathrm{op}}$ and let $\mathcal{B}$ be the class of all absolutely connected topological spaces, i.e., $\mathcal{B} = \{X \in \mathbf{Top}$ such that X cannot be decomposed into any disjoint family $\mathcal{L}$ of nonempty closed subsets with $|\mathcal{L}| > 1\}$ (cf. [Pr₁]). $\mathcal{A}$ and $\mathcal{B}$ are corresponding fixed points of the Galois connection $(\Delta_{\mathcal{N}}, \nabla_{\mathcal{N}})$ of Proposition 5.2. Let $M \xrightarrow{m} X$ be an $\mathcal{M}$-subobject of $X \in \mathbf{Top}$. We have that $M^{S_{\mathbf{Top_1}}} \simeq M^{T_{\mathcal{N}}(\mathbf{Top_1})}$ ([CH, Example 4.3]), i.e., the $T_{\mathcal{N}}(\mathbf{Top_1})$-closure agrees with the Salbany closure induced by $\mathbf{Top_1}$. So, from Theorem 5.7, we have that $\mathcal{B} = I_{\mathcal{N}}(S_{\mathbf{Top_1}})$.

From Proposition 5.6 one can see that $M^{J_{\mathcal{N}}(\mathcal{B})}$ is the union of M with all absolutely connected subsets of X that intersect M. Theorem 5.7 implies that $\mathbf{Top_1} = (D_{\mathcal{N}} \circ J_{\mathcal{N}})(\mathcal{B})$.

We observe that $\mathbf{Top_0}$, $\mathbf{Top_1}$ and all totally disconnected topological spaces are right fixed points of the Galois connection $(D_{\mathcal{N}}, T_{\mathcal{N}})$. Moreover, **Ind**, connected topological spaces and absolutely connected topological spaces are left fixed points of the Galois connection $(J_{\mathcal{N}}, I_{\mathcal{N}})$.

Example 5.9 Let $\mathcal{X}$ be the category **Grp** of groups with the usual (episink,monomorphism)-factorization structure.

(a). Let $\mathcal{N} = \mathcal{M}$ be the class of all monomorphisms in **Grp**. Clearly, to say that a **Grp**-morphism $X \xrightarrow{f} Y$ is $\mathcal{N}$-constant simply means that the image of X under f is a singleton.

Let $\mathcal{A}$ be the subcategory $\mathbf{Ab}$ of abelian groups. We have that $S_{\mathbf{Ab}} \simeq T_{\mathcal{N}}(\mathbf{Ab})$ (cf. [CH, Example 4.4]). $\mathcal{N}$ satisfies the hypotheses of Theorem 5.7 and consequently, $\nabla_{\mathcal{N}}(\mathbf{Ab})$ agrees with $I_{\mathcal{N}}(S_{\mathbf{Ab}})$ which is equal to the class of perfect groups, i.e., $X \in \nabla_{\mathcal{N}}(\mathbf{Ab})$ iff $X = X'$, where X' denotes the subgroup generated by the commutators of X. Thus, $M^{J_{\mathcal{N}}(\nabla(\mathbf{Ab}))}$ is the subgroup generated by M and all perfect subgroups of X and $(D_{\mathcal{N}} \circ J_{\mathcal{N}})(\nabla_{\mathcal{N}}(\mathbf{Ab}))$ is the class of all groups which do not have any non-trivial perfect subgroup.

(b). Let $\mathcal{N}$ be the class of all singleton monomorphisms. Clearly also in this case $\mathcal{N}$-constant simply means constant.

As in part (a), the class $\mathcal{B}$ of perfect groups and the class $\mathcal{A}$ that consists of all groups that do not have any non-trivial perfect subgroup form a pair $(\mathcal{B}, \mathcal{A})$ of corresponding fixed points of $(\Delta_{\mathcal{N}}, \nabla_{\mathcal{N}})$. For every $M \leq X$, $m^{T_{\mathcal{N}}(\mathcal{A})}$ is the intersection of all normal subgroups of X containing M such that $X/M \in \mathcal{A}$. Moreover, $M^{J_{\mathcal{N}}(\mathcal{B})}$ is the subgroup generated by M and all perfect subgroups of X.

Example 5.10 Let $\mathcal{X}$ be the category $\mathbf{Ab}$ of abelian groups with the (episink,monomorphism)-factorization structure.

(a). Let $\mathcal{N} = \mathcal{M}$ be the class of all monomorphisms in $\mathbf{Ab}$. Let $(\mathcal{T}, \mathcal{F})$ be a torsion theory. Clearly, $\mathcal{T}$ and $\mathcal{F}$ are corresponding fixed points of the Galois connection $(\Delta_{\mathcal{N}}, \nabla_{\mathcal{N}})$ of Proposition 5.2. Let $X \in \mathbf{Ab}$ and let $X \xrightarrow{r_X} rX$ be its $\mathcal{F}$-reflection. For every subobject $M \xrightarrow{m} X$ we have that $M^{T_{\mathcal{N}}(\mathcal{F})} \simeq r_X^{-1}(r_X(M)) \simeq M + Ker(r_X)$. Since $\mathcal{T}$ is closed under quotients, $M^{J_{\mathcal{N}}(\mathcal{T})}$ is the subgroup generated by M and all subgroups $S \leq X$ such that $S \in \mathcal{T}$. In particular, if $(\mathcal{T}, \mathcal{F}) =$ (Torsion,Torsion-free), then $M^{T_{\mathcal{N}}(\mathcal{F})} \simeq M + Tor(X)$, where $Tor(X)$ denotes the torsion subgroup of X. If $(\mathcal{T}, \mathcal{F}) =$ (Divisible,Reduced), then $M^{T_{\mathcal{N}}(\mathcal{F})} \simeq M + Div(X)$, where $Div(X)$ denotes the largest divisible subgroup of X. It is interesting to notice that in both cases, $M^{J_{\mathcal{N}}(\mathcal{T})} = M^{T_{\mathcal{N}}(\mathcal{F})}$ (cf. [CH]).

(b). Now let $\mathcal{N}$ be the class of all inclusions of divisible subgroups. Again $\mathcal{N}$-constant means constant. As above, if $(\mathcal{T}, \mathcal{F})$ is a torsion theory, then $\mathcal{T}$ and $\mathcal{F}$ are corresponding fixed points of the Galois connection $(\Delta_{\mathcal{N}}, \nabla_{\mathcal{N}})$ of Proposition 5.2. If $\mathbf{Red}$ is the subcategory of reduced abelian groups, then for every subgroup $M \xrightarrow{m} X$, $M^{T_{\mathcal{N}}(\mathbf{Red})}$ is the intersection of all subgroups of $\mathcal{X}$ containing M such that X/M is reduced. As it is easily seen, this agrees with the Salbany closure $S_{\mathbf{Red}}$. Moreover, if $\mathbf{Div}$ is the subcategory of divisible abelian groups, then for every subgroup $M \xrightarrow{m} X$, $M^{J_{\mathcal{N}}(\mathbf{Div})} \simeq M + Div(X)$.

(c). If $\mathcal{N}$ is the class of all inclusions of torsion subgroups, then also in this case $\mathcal{N}$-constant means constant. If we consider the torsion theory $(\mathbf{T}, \mathbf{Tf})$ where $\mathbf{T}$ is the subcategory of all torsion abelian groups and $\mathbf{Tf}$ is the subcategory of all torsion free abelian groups, then for every subgroup $M \xrightarrow{m} X$, $M^{T_{\mathcal{N}}(\mathbf{Tf})}$ is the intersection of all subgroups of $\mathcal{X}$ containing M such that X/M is torsion free. As it is easily seen, this agrees with the

Salbany closure $S_{\mathbf{Tf}}$. Moreover, for every subgroup $M \xrightarrow{m} X$, $M^{J_{\mathcal{N}}(\mathbf{T})} \simeq M + Tor(X)$.

Remark 5.11 It is important to observe that a careful look at the above examples already reveals that the real importance of the notion of $\mathcal{N}$-constant morphism does not lie in the fact that it provides a general notion of constant morphism in an arbitrary category but rather in the fact that it provides an alternative way of describing the composition of the two Galois connections $S(\mathcal{X}) \underset{I_{\mathcal{N}}}{\overset{J_{\mathcal{N}}}{\rightleftarrows}} CL(\mathcal{X},\mathcal{M})$ and $CL(\mathcal{X},\mathcal{M}) \underset{T_{\mathcal{N}}}{\overset{D_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathbf{op}}$. Clearly this role is particularly fulfilled whenever in concrete situations $\mathcal{N}$-constant means constant in the classical sense. This point of view will become more apparent as we go further into the theory.

Despite the fact that this theory seemed to be quite successful, the requirement that $\mathcal{N}$ be closed under direct images created some sort of limitation to the range of applications of Theorem 5.7. To go around this problem, the following new definition was introduced in [C7].

Definition 5.12 A $\mathcal{X}$-morphism $X \xrightarrow{f} Y$ is called $\mathcal{N}$-fixed if for every $n \in \mathcal{N}_Y$ we have that $f^{-1}(n) \simeq id_X$.

By replacing $\mathcal{N}$-constant with $\mathcal{N}$-fixed in Proposition 5.2, one obtains a new Galois connection $S(\mathcal{X}) \underset{\hat{\nabla}_{\mathcal{N}}}{\overset{\hat{\Delta}_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathbf{op}}$.

The closure properties of the Galois fixed points of this Galois connection were studied in [C7] and the following result was obtained.

Theorem 5.13 *Let $\mathcal{N}$ be a subclass of $\mathcal{M}$ closed under the formation of pullbacks. Then we have that the Galois connection $S(\mathcal{X}) \underset{\hat{\nabla}_{\mathcal{N}}}{\overset{\hat{\Delta}_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathbf{op}}$ factors through $CL(\mathcal{X},\mathcal{M})$ via the Galois connections $S(\mathcal{X}) \underset{I_{\mathcal{N}}}{\overset{J_{\mathcal{N}}}{\rightleftarrows}} CL(\mathcal{X},\mathcal{M})$ and $CL(\mathcal{X},\mathcal{M}) \underset{T_{\mathcal{N}}}{\overset{D_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathbf{op}}$.*

Notice that there is a strong relationship between the definition of $\mathcal{N}$-constant morphism that appeared in [C5] and the one of $\mathcal{N}$-fixed morphism in [C7], as the following proposition shows.

Proposition 5.14 *Let $X \xrightarrow{f} Y$ be a $\mathcal{X}$-morphism. Then we have the following:*

(a) If $\mathcal{N}$ is closed under the formation of pullbacks, then every $\mathcal{N}$-constant morphism is $\mathcal{N}$-fixed;

(b) If $\mathcal{N}$ is closed under the formation of direct images, then every $\mathcal{N}$-fixed morphism is $\mathcal{N}$-constant;

(c) If $\mathcal{N}$ is closed under the formation of both pullbacks and direct images, then the two notions of $\mathcal{N}$-constant and $\mathcal{N}$-fixed morphism are equivalent.

Conditions under which the closure operator $J_{\mathcal{N}}(\mathcal{B})$ is idempotent and the closure operator $T_{\mathcal{N}}(\mathcal{A})$ is weakly hereditary are also studied in [C$_7$].

The next example, whose details can be found in [C$_7$], shows the importance of the definition of $\mathcal{N}$-fixed morphism.

Example 5.15 Let $\mathcal{X}$ be the category **Grp** of groups with the usual (episink,monomorphism)-factorization structure and let $\mathcal{N}$ consist of all inclusions of normal subgroups. Clearly $\mathcal{N}$ is closed under the formation of pullbacks but not under the formation of direct images, so Theorem 5.13 applies but 5.7 does not. Notice that in this case, $X \xrightarrow{f} Y$ is $\mathcal{N}$-fixed if and only if f is constant in the classical sense.

Let **Sim** denote the subcategory of simple groups, i.e., all those groups that have no nontrivial normal subgroups. Then we have that $\hat{\Delta}_{\mathcal{N}}(\mathbf{Sim}) =$ **Simfree**, i.e., the subcategory of all groups that have no simple subgroup different from zero.

Moreover, we have that $\hat{\nabla}_{\mathcal{N}}(\mathbf{Simfree}) = \mathbf{Simquo}$, i.e., the subcategory of all groups X such that if K is a normal subgroup of X, then X/K has a simple subgroup different from zero.

It is important to observe that the above example shows the usefulness of being able to describe the connectedness-disconnectedness Galois connection by means of the notion of $\mathcal{N}$-fixed morphism. As a matter of fact it is quite difficult to characterize the Galois closed classes, using Corollary 5.5 directly. The problem lies in the fact that it is not easy to characterize $T_{\mathcal{N}}(\mathbf{Simfree})$. Notice that the previous notion of $\mathcal{N}$-constant morphism cannot be used in this case since $\mathcal{N}$ is not closed under the formation of direct images.

At this point, comparing the results in Theorems 5.7 and 5.13, it became clear that two different descriptions of the Galois connection $S(\mathcal{X}) \underset{I_{\mathcal{N}} \circ T_{\mathcal{N}}}{\overset{D_{\mathcal{N}} \circ J_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ had been obtained, depending on the chosen assumptions on the subclass $\mathcal{N}$. This naturally lead to the conclusion that this Galois connection could be the appropriate tool to introduce a notion of connectedness in an arbitrary category, and that the two notions of $\mathcal{N}$-constant and $\mathcal{N}$-fixed morphism could be used to provide additional descriptions of it, depending on whether the class $\mathcal{N}$ is closed under the formation of direct images or under the formation of pullbacks.

Therefore in [C$_8$] the following definition was given.

Definition 5.16 A $\mathcal{X}$-object X is called $(C,\mathcal{N})$-connected if $X \in I_{\mathcal{N}}(T_{\mathcal{N}}(D_{\mathcal{N}}(C)))$.

As a consequence, the theorems mentioned above can be used to provide alternative descriptions of the notion of $(C,\mathcal{N})$-connectedness under the appropriate closedness conditions as follows.

Proposition 5.17

(a) If $\mathcal{N}$ is closed under the formation of direct images, then an $\mathcal{X}$-object X is $(C,\mathcal{N})$-connected if every morphism $X \xrightarrow{f} A$ with $A \in D_{\mathcal{N}}(C)$ is $\mathcal{N}$-constant; i.e., f factors through n_f for every $n \in \mathcal{N}_X$.

(b) If $\mathcal{N}$ is closed under the formation of pullbacks, then an $\mathcal{X}$-object X is $(C,\mathcal{N})$-connected if every morphism $X \xrightarrow{f} A$ with $A \in D_{\mathcal{N}}(C)$ is $\mathcal{N}$-fixed; i.e., $f^{-1}(n) \simeq id_X$ for every $n \in \mathcal{N}_A$.

The following generalizations of classical topological results hold.

Proposition 5.18 *Let $\mathcal{N}$ be closed under the formation of pullbacks along morphisms in $\mathbf{E}$.*

(a) If $X \xrightarrow{f} Y$ belongs to $\mathbf{E}$ and X is $(C,\mathcal{N})$-connected, then so is Y.

(b) Let $X \xrightarrow{f} Y$ be a $\mathcal{X}$-morphism. If X is $(C,\mathcal{N})$-connected, so is X_f.

If moreover $\mathcal{N}$ is closed under the formation of all pullbacks then we have:

(c) If $(X_i)_{i\in I}$ is a family of $(C,\mathcal{N})$-connected $\mathcal{X}$-objects and the coproduct $\amalg X_i$ exists, then it is $(C,\mathcal{N})$-connected.

Remark 5.19 Suppose that the category $\mathcal{X}$ has products and that each projection belongs to $\mathbf{E}$. Moreover, assume that $\mathcal{N}$ is closed under the formation of pullbacks along morphisms in $\mathbf{E}$. Then from Proposition 5.18 (a) we obtain that if the product of a family of $\mathcal{X}$-objects is $(C,\mathcal{N})$-connected, so is each of its factors. However, the converse is not true. As a counterexample, it is enough to consider in the category $\mathbf{Ab}$ of abelian groups, the subcategory $\mathbf{T}$ consisting of all torsion abelian groups. Clearly, $\mathbf{Ab}$ satisfies our assumptions. As Example 5.10 (a) shows, the subcategory $\mathbf{T}$ is the connectedness class of a certain closure operator, but it is not closed under products.

Remark 5.20 It may be interesting to observe that in the case that $\mathcal{X}$ is well-powered and has coproducts, if $\mathcal{N}$ is closed under the formation of pullbacks, (b) and (c) of Proposition 5.18 imply that for any closure operator C, the $(C,\mathcal{N})$-connected objects form an $\mathcal{M}$-coreflective subcategory of $\mathcal{X}$ (cf. [AHS, Theorem 16.8], dual).

We also have the following:

Proposition 5.21 *Let $\mathcal{N}$ be closed under the formation of pullbacks along morphisms in $\mathcal{M}$.*

(a) Let $M \xrightarrow{m} X$ be a C-dense $\mathcal{M}$-subobject of $X \in \mathcal{X}$. If M is $(C,\mathcal{N})$-connected, then so is X.

(b) Let C be weakly hereditary and let $M \xrightarrow{m} X$ be an $\mathcal{M}$-subobject. If M is $(C,\mathcal{N})$-connected then so is M^{C}.

(c) Let $(M_i \xrightarrow{m_i} X)_{i\in I}$ be a family of $\mathcal{M}$-subobjects of $X \in \mathcal{X}$. If each M_i is $(C,\mathcal{N})$-connected then so is their supremum $\vee M_i$.

Further generalizations of classical results about topological connectedness can be found in [C$_8$] under the assumption that $\mathcal{N}$ is closed under direct images and $\mathcal{X}$ contains a terminal object T. For instance the notion of C-component was introduced and some related results were proved. We encourage the reader who has further interest in this topic to check [C$_8$].

The limitations introduced by the closedness requirements of the subclass $\mathcal{N}$ in Theorems 5.7 and 5.13 were finally removed in [C$_9$] by means of the introduction of a new notion of constant morphism that was termed $\mathcal{N}$-*dependent.* We briefly recall this achievement.

Definition 5.22 A morphism $X \xrightarrow{f} Y$ is $\mathcal{N}$-dependent if for every $n \in \mathcal{N}_X$ and every $p \in \mathcal{N}_Y$, $n_f \leq p$ implies $f^{-1}(p) \simeq id_X$.

This notion is strongly related to the ones of $\mathcal{N}$-constant and $\mathcal{N}$-fixed morphism as the following proposition shows:

Proposition 5.23 *For a morphism $X \xrightarrow{f} Y$ consider the statements:*

(a) f is $\mathcal{N}$-dependent;

(b) f is $\mathcal{N}$-constant;

(c) f is $\mathcal{N}$-fixed.

We always have that (b) $\Rightarrow$ (a) and (c) $\Rightarrow$ (a). If $\mathcal{N}$ is closed under the formation of direct images, then (a) $\Leftrightarrow$ (b) $\Leftarrow$ (c). If $\mathcal{N}$ is closed under the formation of pullbacks, then (a) $\Leftrightarrow$ (c) $\Leftarrow$ (b). As a consequence, if $\mathcal{N}$ is closed under the formation of both pullbacks and direct images, then the three concepts are equivalent.

Clearly, the notion of $\mathcal{N}$-dependent morphism yields a Galois connection $S(\mathcal{X}) \underset{\nabla'_{\mathcal{N}}}{\overset{\Delta'_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ where for $\mathcal{A} \in S(\mathcal{X})$, $\Delta'_{\mathcal{N}}(\mathcal{A}) = \{Y \in \mathcal{X} : \forall X \in \mathcal{A}, X \xrightarrow{f} Y$ is $\mathcal{N}$-dependent$\}$ and for $\mathcal{B} \in S(\mathcal{X})^{\mathrm{op}}$, $\nabla'_{\mathcal{N}}(\mathcal{B}) = \{X \in \mathcal{X} : \forall Y \in \mathcal{B}, X \xrightarrow{f} Y$ is $\mathcal{N}$-dependent$\}$. Thus we obtain the following:

Theorem 5.24 *The Galois connection* $S(\mathcal{X}) \underset{\nabla'_{\mathcal{N}}}{\overset{\Delta'_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ *factors through* $CL(\mathcal{X},\mathcal{M})$ *via the Galois connections* $S(\mathcal{X}) \underset{I_{\mathcal{N}}}{\overset{J_{\mathcal{N}}}{\rightleftarrows}} CL(\mathcal{X},\mathcal{M})$ *and* $CL(\mathcal{X},\mathcal{M}) \underset{T_{\mathcal{N}}}{\overset{D_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$.

The rest of [C$_9$] is devoted to the task of obtaining a characterization of C-connectedness classes, that is, classes of the form $I_{\mathcal{N}}(T_{\mathcal{N}}(D_{\mathcal{N}}(C)))$. As a matter of fact, under some mild additional assumptions on $\mathcal{X}$, among which the existence of a terminal object, a characterization of C-connectedness classes that closely resembles the one in [AW] is obtained. A similar characterization for C-disconnectedness classes is in the works (cf. [C$_{10}$]).

Next we present a different categorical approach to connectedness with respect to a closure operator due to Clementino and Tholen [CT$_1$].

The setup used is the one of a finitely complete $(\mathcal{E},\mathcal{M})$-category $\mathcal{X}$ in which $\mathcal{E}$ is a class of epimorphisms and $\mathcal{M}$ is a class of monomorphisms, both containing the $\mathcal{X}$-isomorphisms. It is also assumed that $\mathcal{X}$ is $\mathcal{M}$-complete so that $\mathcal{X}$ has multiple pullbacks of arbitrary sinks of $\mathcal{M}$-morphisms. It is well know that in this case, the above factorization structure for morphisms $(\mathcal{E},\mathcal{M})$ extends to a factorization structure for sinks $(\mathbf{E},\mathcal{M})$.

Unfortunately the results that we wish to present here require a fairly high number of new concepts besides the ones already introduced in Section 3. Here we will recall only those that are absolutely necessary for the understanding of the included results.

In [CT$_1$] the category $\mathcal{X}$ was said to *have enough quasipoints* (*points*) if for every $X \in \mathcal{X}$ the supremum of all its quasipoints (points) is isomorphic to 1_X.

We recall that a closure operator C is called *finitely productive* if for any pair of $\mathcal{M}$-subobjects $M \xrightarrow{m} X$ and $N \xrightarrow{n} Y$, $\overset{C}{(m\times n)}_{X\times Y} \simeq \overset{C}{(m)}_X \times \overset{C}{(n)}_Y$ holds.

A full subcategory $\mathcal{A}$ of $\mathcal{X}$ is *upwards closed* if every epimorphism $X \xrightarrow{e} Y$ satisfies the property: if for every $Y \in \mathcal{A}$ and for every quasipoint $y \xrightarrow{i_y} Y$, the pullback $e^{-1}(y)$ belongs to $\mathcal{A}$, then so does X. If in this definition we replace epimorphism with regular epimorphism, then the class $\mathcal{A}$ is called *q-reversible.* $\mathcal{A}$ is called *second-additive* if it is closed under $\mathbf{E}$-sinks with a common quasipoint.

$\mathcal{A}$ is closed under C-dense extensions if for every C-dense $\mathcal{M}$-subobject $M \xrightarrow{m} X$, $M \in \mathcal{A}$ implies $X \in \mathcal{A}$.

Definition 5.25 A $\mathcal{X}$-object X is called *C-connected* if the diagonal morphism $X \xrightarrow{\delta_X} X \times X$ is C-dense.

The full subcategory of all C-connected objects will be denoted by $\nabla(C)$.

One obtains the following properties of $\nabla(C)$. If C is (finitely) productive, then $\nabla(C)$ is closed under (finite) products in $\mathcal{X}$. Under mild hypotheses, it is also closed under $\mathcal{E}$-images, C-dense extensions and C-dense subobjects.

The following result was obtained.

Theorem 5.26 *Let $\mathcal{X}$ have enough quasipoints. Then a full subcategory $\mathcal{A}$ of $\mathcal{X}$ is of the form $\nabla(C)$ for some C if and only if $\mathcal{A}$ contains all preterminal objects and the following condition holds:*

for every sink $(A_i^2 \xrightarrow{h_i} X^2)_{i\in I}$ with $A_i \in \mathcal{A}$ and $(\delta_{A_i})_{h_i} \leq \delta_X$ for all $i \in I$ and $1_{X^2} \simeq \bigvee_{i\in I}(1_{A_i^2})_{h_i}$, one has $X \in \mathcal{A}$.

Definition 5.27 A morphism $X \xrightarrow{f} Y$ is *constant* if its $\mathcal{E}$-image is preterminal.

Definition 5.28 For subcategories $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{X}$, the *left* and *right-constant* subcategories of $\mathcal{A}$ and $\mathcal{B}$ are defined respectively by:

$l(\mathcal{A}) = \{X \in \mathcal{X} : \forall Y \in \mathcal{A}, X \xrightarrow{f} Y \text{ is constant}\}$,

$r(\mathcal{B}) = \{Y \in \mathcal{X} : \forall X \in \mathcal{B}, X \xrightarrow{f} Y \text{ is constant}\}$.

This defines a Galois connection whose fixed points are called left and right constant (cf. [H]).

A right constant subcategory is trivially closed under monosources. It therefore contains all preterminal objects and is closed under limits in $\mathcal{X}$; it is even strongly epireflective if $\mathcal{X}$ has products and is $\mathcal{E}$-co-well powered. A left constant category is closed under $\mathcal{E}$-images.

The following two characterizations of right and left constant subcategories are presented in [CT$_1$]. Their purpose is to provide a generalization of the topological analogues that appeared in [AW]. Each of the following results comes with a few additional hypotheses on the category $\mathcal{X}$ that would require additional definitions. That would be beyond the scope of this paper. Therefore, we strongly recommend the reader interested in this topic to check the additional requirements directly from [CT$_1$].

Theorem 5.29 *Let C be a closure operator on $\mathcal{X}$ such that the quasipoints are C-closed and strong epimorphisms are stable under pullbacks along C-closed subobjects. Then a full reflective subcategory $\mathcal{A}$ of $\mathcal{X}$ is right constant if and only if*

(a) $\mathcal{A}$ is closed under monomorphisms,

(b) $\mathcal{A}$ is upwards closed.

For a second-additive subcategory $\mathcal{A}$ that contains all preterminal objects, one can define the *$\mathcal{A}$-component* of a quasipoint $x \xrightarrow{i_x} X$ as the largest subobject of $A_x \xrightarrow{a_x} X$ such that $i_x \leq a_x$ and $A_x \in \mathcal{A}$. Then we say that a second-additive subcategory $\mathcal{A}$ that contains all preterminal objects is called *strongly second-additive* if every $\mathcal{X}$-object X for which the $\mathcal{A}$-component of each of its quasipoints x is isomorphic to x, belongs to $r(\mathcal{A})$.

Theorem 5.30 *Let C be a closure operator on $\mathcal{X}$ such that the regular epimorphisms are stable under pullbacks along C-closed subobjects, and that every C-closed subobject is a pullback of a quasipoint. Then a full subcategory $\mathcal{B}$ of $\mathcal{X}$ which is closed under C-dense extensions is left constant if and only if:*

(a) $\mathcal{B}$ is closed under $\mathcal{E}$-images,

(b) $\mathcal{B}$ is strongly second-additive,

(c) $\mathcal{B}$ is q-reversible.

We conclude with the following interesting result:

Theorem 5.31 *Let $\mathcal{X}$ be $\mathcal{E}$-co-well powered with products and enough quasipoints. Then, for full subcategories $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{X}$ we have:*

(a) $\mathcal{B}$ is left constant if and only if $\mathcal{B} = \nabla(C)$ for some regular closure operator C on $\mathcal{X}$.

(b) $\mathcal{A}$ is right constant if and only if $\mathcal{A} = \Delta(C)$ for some coregular closure operator C on $\mathcal{X}$.

The definition of coregular closure operator can be found in [CT_1]. Moreover, some interesting diagrams that further illustrate the relationship between closure operators and left and right constant subcategories can be also found there. We do not include them here for space reasons. We conclude the section dedicated to [CT_1] with some examples that illustrate the notions of C-connectedness and C-separation introduced in that paper.

Example 5.32 Consider the category **Top** of topological spaces with the (surjective,embedding)-factorization structure.

(a). If K is the Kuratowski closure, that is the closure induced by the topology, then $\Delta(K)$ is the subcategory of Hausdorff spaces and $\nabla(K)$ consists of the irreducible spaces, i.e., those spaces whose disjoint open subsets U, V must satisfy either $U = \emptyset$ or $V = \emptyset$.

(b). If Θ is the Θ-closure, then $\Delta(\Theta)$ is the subcategory of Urysohn spaces, i.e., those spaces in which distinct points can be separated by disjoint closed neighborhoods and $\nabla(\Theta) = \{X \in \mathbf{Top} :$ if $U \subseteq X, V \subseteq X,$ are open subsets of X with $\bar{U} \cap \bar{V} = \emptyset,$ then $U = \emptyset$ or $V = \emptyset\}$.

(c). Let C be the closure operator that to each $M \subseteq X$ associates the intersection of all clopen subsets of X containing M. Then, $\Delta(C)$ is the subcategory of totally disconnected spaces and $\nabla(C)$ consists of all connected topological spaces.

(d). For the sequential closure Σ, $\Delta(\Sigma)$ is the subcategory of $\mathcal{US}$ spaces, i.e., all those spaces in which convergent sequences have uniquely determined limits and $\nabla(\Sigma)$ consists of all those spaces X such that for each pair of points $x, y \in X$, there is a sequence converging to both x and y.

Consider the category **Grp** with its (epimorphism,monomorphism)-factorization structure.

(e). Let C^ν be the normal closure operator. Then $\Delta(C^\nu)$ is the subcategory of abelian groups and $\nabla(C^\nu)$ consists of all those groups G such that for all $x, y \in G$ there are $z_1, \ldots, z_n \in G$ such that x, y both belong to the set $c_G(z_1) \cdot c_G(z_2) \cdot \cdots \cdot c_G(z_n)$, where $c_G(z)$ is the conjugacy class of $z \in G$.

(f). For every group G, let $k(G)$ denote the commutator subgroup of G and let C^k be the closure operator on **Grp** defined by: $C^k_G(M) = k(G) \cdot M$, for every subgroup M of G. Then, $\Delta(C^k)$ is the subcategory of abelian groups and $\nabla(C^k)$ is the subcategory of perfect groups, i.e., those groups that coincide with their commutator subgroup.

We conclude this section with the main results on connectedness via closure operators that appeared in [Cl$_5$].

The setup of this paper is the same as in [CT$_1$]. The idea of the author is to generalize the main result in [CH] to an arbitrary category $\mathcal{X}$. This generalization should be compatible with the notions of left and right-constant subcategories that appear in Definition 5.28.

As in [CT$_1$], $\mathcal{P}$ denotes the full subcategory of all preterminal objects.

For an $\mathcal{M}$-subobject $M \xrightarrow{m} X$ the following two closure operators are defined:

$$fine_X(m) = m \quad \text{and} \quad coar_X(m) = \wedge\{f^{-1}(f(m)) | X \xrightarrow{f} P, P \in \mathcal{P}\}.$$

Consequently, for a closure operator C, the following full subcategories are defined:

$$\begin{aligned} Coar(C) &= \{A \in \mathcal{X} | C_A \geq coar_A\} \quad \text{and} \\ Fine(C) &= \{B \in \mathcal{X} | C_B \leq fine_B\}. \end{aligned}$$

Moreover, for subcategories $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{X}$, the closure operators $coar^{\mathcal{B}}$ and $fine^{\mathcal{A}}$ are defined:

$$coar^{\mathcal{B}}_X(m) = \wedge\{f^{-1}(f(m)) | f : X \longrightarrow B, B \in \mathcal{B}\} \qquad \text{and}$$
$$fine^{\mathcal{A}}_X(m) = m \vee \sup\{h(coar_A(h^{-1}(m))) | h : A \longrightarrow X, A \in \mathcal{A}\}.$$

In conclusion, the following two Galois connections are obtained:

$$S(\mathcal{X}) \underset{Coar}{\overset{fine}{\rightleftarrows}} CL(\mathcal{X}, \mathcal{M}) \qquad \text{and} \qquad CL(\mathcal{X}, \mathcal{M}) \underset{coar}{\overset{Fine}{\rightleftarrows}} S(\mathcal{X})^{\mathbf{op}}.$$

The following main result that links the above Galois connections to the one in Definition 5.28 is proved.

Theorem 5.33 *If $\mathcal{E}$ is stable under pullbacks along monomorphisms and $\mathcal{P}$ is closed under images , then the Galois connection $S(\mathcal{X}) \underset{l}{\overset{r}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ factors through $CL(\mathcal{X},\mathcal{M})$ via the Galois connections $S(\mathcal{X}) \underset{Coar}{\overset{fine}{\rightleftarrows}} CL(\mathcal{X},\mathcal{M})$ and $CL(\mathcal{X},\mathcal{M}) \underset{coar}{\overset{Fine}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ if and only if $\mathcal{P} = \{X | coar_X = fine_X\}$.*

We would like to observe that the Galois connection $CL(\mathcal{X},\mathcal{M}) \underset{coar}{\overset{Fine}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ is a special case of the Galois connection $CL(\mathcal{X},\mathcal{M}) \underset{T_{\mathcal{N}}}{\overset{D_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ in the case $\mathcal{N}$=$\mathcal{M}$ (cf. Proposition 5.3). However, the Galois connection $S(\mathcal{X}) \underset{Coar}{\overset{fine}{\rightleftarrows}} CL(\mathcal{X},\mathcal{M})$ is quite different from the Galois connection $S(\mathcal{X}) \underset{I_{\mathcal{N}}}{\overset{J_{\mathcal{N}}}{\rightleftarrows}} CL(\mathcal{X},\mathcal{M})$ obviously because the Galois connection $S(\mathcal{X}) \underset{l}{\overset{r}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$ is in general different from $S(\mathcal{X}) \underset{\nabla_{\mathcal{N}}}{\overset{\Delta_{\mathcal{N}}}{\rightleftarrows}} S(\mathcal{X})^{\mathrm{op}}$.

The paper ends with two interesting results followed by some examples. In order to present these results we need to clarify some terminology.

One says that *pairs of subobjects detect monosources* if any source $(X \xrightarrow{f_i} Y_i)_{i\in I}$ is monic, provided that for every pair of $\mathcal{M}$-subobjects $x, y : Y \longrightarrow X$, $f_i \circ x = f_i \circ y$, for every $i \in I$ implies $x = y$. Moreover, a sink $(Y_i \xrightarrow{g_i} Y)_{i\in I}$ is pt-*surjective* if for each $T \xrightarrow{y} Y$ there are $j \in J$ and $T \xrightarrow{y_j} Y_j$ such that $g_j \circ y_j = y$.

We conclude with the following:

Theorem 5.34 *Let C be a closure operator.*

(a) If pairs of subobjects detect monosources, then every subcategory of the form $\Delta(C)$ is a subcategory of type $Fine(D)$ for a suitable closure operator D.

(b) If $\mathcal{X}$ has enough points and sinks in $\mathbf{E}$ are pt-surjective, then every subcategory of the form $\nabla(C)$ is a subcategory of type $Coar(D)$ for some closure operator D.

We conclude by reminding the reader that the results presented in this article are only a small portion of those that actually appeared in the mentioned papers. So, once more we strongly encourage the readers interested in any specific topic mentioned here to refer to the related papers for a deeper insight into that topic.

As a final remark, we would like to emphasize that the theory and applications of categorical closure operators are still being actively worked on. For the latest work on this subject, the interested reader could check the following papers: [BG], [C_{9-10}], [CGT_2], [CT_3], [GT].

References

AC_1 I.W. Alderton and G. Castellini. Epimorphisms in categories of separated fuzzy topological spaces, *Fuzzy Sets and Systems*, **56** (1993), 323–330.

AC_2 I.W. Alderton and G. Castellini. Closure operators and fuzzy connectedness, *Fuzzy Sets and Systems*, **104(3)** (1999), 397–405.

AHS J. Adámek, H. Herrlich and G.E. Strecker. *Abstract and Concrete Categories*, Wiley, New York, 1990.

AW A.V. Arhangel'skii and R. Wiegandt. Connectedness and disconnectedness in topology, *Gen. Top. Appl.*, **5** (1975) 9–33.

B S. Baron. A note on epi in T_0, *Canad. Math. Bull.*, **11** (1968), 503–504.

BG G.C.L. Brümmer and E. Giuli. Splitting closure operators, preprint.

C_1 G. Castellini. Closure operators, monomorphisms and epimorphisms in categories of groups, *Cahiers Topologie Geom. Differentielle Categoriques*, **27(2)** (1986), 151–167.

C_2 G. Castellini. Closure operators and functorial topologies, *J. Pure Appl. Algebra*, **55** (1988), 251–259.

C_3 G. Castellini. Compact objects, surjectivity of epimorphisms and compactifications, *Cahiers Topologie Geom. Differentielle Categoriques*, **31(1)** (1990), 53–65.

C_4 G. Castellini. Regular closure operators and compactness, *Cahiers Topologie Geom. Differentielle Categoriques*, **33(1)** (1992), 21–31.

C_5 G. Castellini. Connectedness, disconnectedness and closure operators, a more general approach, *Proceedings of the workshop on categorical topology*, (L'Aquila, 1994), Kluwer Academic Publisher (1996), 129–138.

C_6 G. Castellini. Connectedness, disconnectedness and closure operators, further results, *Quaestiones Mathematicae*, **20(4)** (1997), 611–638.

C_7 G. Castellini. Connectedness and disconnectedness: a different perspective, *Quaestiones Mathematicae*, **21(1&2)** (1998), 117–133.

C_8 G. Castellini. Connectedness with respect to a closure operator, *Journal of Applied Categorical Structures*, to appear.

C_9 G. Castellini. Connectedness classes, *Quaestiones Mathematicae*, **23(3)** (2000), 313–334.

C_{10} G. Castellini. Disconnectedness classes, *Quaestiones Mathematicae*, to appear.

CC_1 F. Cagliari and M. Cicchese. Epireflective subcategories and epiclosure, *Riv. Mat. Univ. Parma*, **8** (1982) 1–8.

CC_2 F. Cagliari and M. Cicchese. Epireflective subcategories and semiclosure operators, *Quaestiones Mathematicae*, **6** (1983) 295–301.

CC_3 F. Cagliari and M. Cicchese. Disconnectednesses and closure operators, *Proceedings of the 13th Winter School on Abstract Analysis, Topology Section, Supplementi ai Rendiconti del Circolo Matematico di Palermo*, Serie II, **11** (1985) 15–23.

CG_1 G. Castellini and E. Giuli. Closure operators and injectivity, *Comment. Math. Univ. Carol.*, **33(1)** (1992), 149–157.

CG_2 G. Castellini and E. Giuli. Closure operators with respect to a functor, *Journal of Applied Categorial Structures*, to appear.

CGT_1 M. M. Clementino, E. Giuli and W. Tholen. Topology in a category: compactness, *Port. Math.*, **53** (1996), 397–433.

CGT_2 M. M. Clementino, E. Giuli and W. Tholen. What is a quotient map with respect to a closure operator?, preprint.

CH G. Castellini and D. Hajek. Closure operators and connectedness, *Topology and its Appl.*, **55** (1994), 29–45.

CKS_1 G. Castellini, J. Koslowski, and G.E. Strecker. A factorization of the Pumplün–Röhrl connection, *Topology and its Appl.*, **44** (1992), 69–76.

CKS_2 G. Castellini, J. Koslowski and G.E. Strecker. Hereditary and modal closure operators, *Proceedings of the Summer International Meeting*, Montreal, Canada, June 1991), C.M.S. Conference Proceedings, AMS (1992) 111–132.

CKS$_3$ G. Castellini, J. Koslowski and G.E. Strecker. Closure operators and polarities, *Proceedings of the 1991 Summer Conference on General Topology and Applications in Honor of Mary Ellen Rudin and Her Work*, Annals of the New York Academy of Sciences, Vol. 704 (1993), 38–52.

CKS$_4$ G. Castellini, J. Koslowski and G.E. Strecker, Regular closure operators, *Journal of Applied Categorical Structures*, **2** (1994), 219–244.

CKS$_5$ G. Castellini, J. Koslowski and G.E. Strecker. An approach to a dual of regular closure operators, *Cahiers Topologie Geom. Differentielle Categoriques*, **35(2)** (1994), 109–128.

Cl$_1$ M. M. Clementino. Separação e Compacidad em Categorias, Ph.D. Thesis, Universidade de Coimbra, 1992.

Cl$_2$ M. M. Clementino. Weakly hereditary regular closure operators, *Topology and its Appl.*, **49** (1993), 129–139.

Cl$_3$ M. M. Clementino. $\mathcal{T}_0$ and $\mathcal{T}_1$-reflections, *Comment. Math. Univ. Carol.*, **33(3)** (1992), 493–504.

Cl$_4$ M. M. Clementino. On categorical notions of compact objects, *Appl. Categorical Structures*, **4** (1996), 15–29.

Cl$_5$ M. M. Clementino. On connectedness via closure operators, preprint.

CS G. Castellini and G.E. Strecker. Closure operators versus subcategories, *Quaestiones Mathematicae*, **13** (1990), 417–424.

CT$_1$ M. M. Clementino and W. Tholen. Separation versus connectedness, *Topology and its Appl.*, **75** (1997), 143–181.

CT$_2$ M. M. Clementino and W. Tholen. Tychonoff's theorem in a category, *Proc. Amer. Math. Soc.*, **124** (1996), 3311–3314.

CT$_3$ M. M. Clementino and W. Tholen. Separated and connected maps, preprint.

D D. Dikranjan. Semiregular closure operators and epimorphisms in topological categories, *Rend. Circ. Mat. Palermo*, Serie II, **29** (1992), 105–160.

DG$_1$ D. Dikranjan and E. Giuli. Closure operators induced by topological epireflections, *Coll. Math. Soc. J. Bolyai*, **41** (1983), 233–246.

DG_2 D. Dikranjan and E. Giuli. Epimorphisms and cowellpoweredness of epireflective subcategories of **Top**, *Rend. Circ. Mat. Palermo*, **(2)** Suppl. no. 6 (1984), 121–136.

DG_3 D. Dikranjan and E. Giuli. Closure operators I, *Topology and its Appl.*, **27** (1987), 129–143.

DG_4 D. Dikranjan and E. Giuli. Factorizations, injectivity and compactness in categories of modules, *Comm. Algebra*, **19(1)** (1991), 45–83.

DG_5 D. Dikranjan and E. Giuli. Compactness, minimality and closedness with respect to a closure operator, *Categorical Topology and its Relation to Analysis Algebra and Combinatorics*, (Conference Proceeding, Prague 1998), World Scientific, Singapore (1989), 284–296.

DG_6 D. Dikranjan and E. Giuli. C-perfect morphisms and C-compactness, unpublished manuscript.

DG_7 D. Dikranjan and E. Giuli. Epis in categories of convergence spaces, *Acta Math. Hungar.*, **61** (1993), 195–201.

DG_8 D. Dikranjan and E. Giuli. Uryshon-closed spaces - old and new, unpublished manuscript.

DG_8 D. Dikranjan, E. Giuli. S(n)-θ-closed spaces, *Topology and its Appl.*, **28** (1988), 59–74.

DGT D. Dikranjan, E. Giuli and W. Tholen. Closure operators II, *Proceedings of the Conference in Categorical Topology*, (Prague, 1988), World Scientific (1989), 297–335.

DGTo D. Dikranjan, E. Giuli and A. Tozzi. Topological categories and closure operators, *Quaestiones Mathematicae*, **11** (1988), 323–337.

DT D. Dikranjan and W. Tholen. *Categorical Structure of Closure Operators, with Applications to Topology, Algebra and Discrete Mathematics*, Kluwer Academic Publishers, 1995.

DW D. Dikranjan and S. Watson. The category of $S(\alpha)$-spaces is not cowellpowered, preprint.

EKMS M. Erné, J. Koslowski, A. Melton and G. Strecker. A primer on Galois connections, *Proceedings of the 1991 Summer Conference on General Topology and Applications in Honor of Mary Ellen Rudin and Her Work*, Annals of the New York Academy of Sciences, Vol. 704 (1993), 103–125.

F_1 T. H. Fay. Compact modules, *Comm. Alg.*, **16** (1988), 1209–1219.

F_2 T. H. Fay. Divisible, compact and cotorsion modules, preprint.

F_3 T. H. Fay. Weakly hereditary regular closure operators, *Cahiers Topologie Geom. Differentielle Categoriques*, **37(4)** (1996), 279–293.

Fe_1 A. Fedeli. $\mathcal{A}$-locally compact spaces, *Rend. Istit. Mat. Univ. Trieste*, **20** (1988), 175–185.

Fe_2 A. Fedeli. Operatori di chiusura e spazi $T(\mathcal{A})$, *Atti Sem. Mat. Fis. Univ. Modena*, **37** (1989), 45–51.

Fe_3 A. Fedeli. Two properties related to $\mathcal{A}$-compactness, *Riv. Mat. Univ. Parma*, **15(4)** (1989), 89–98.

FJ T.H. Fay and S.V. Joubert. Categorical compactness for rings, *J. Austr. Math. Soc. Ser. A*, **59** (1995), 313–329.

FS T.H. Fay and M.J. Schoeman. The Z-adic completion, the p-adic completion and categorical compactness for abelian groups, preprint.

FW_1 T.H. Fay and G.L. Walls. Regular and normal closure operators and categorical compactness for groups, *Journal of Applied Categorical Structures*, **3** (1995), 261–278.

FW_2 T.H. Fay and G.L. Walls. Completions and categorical compactness for nilpotent groups, *Quaestiones Mathematicae*, **17(4)** (1994), 437–451.

FW_3 T.H. Fay and G.L. Walls. Categorically compact locally nilpotent groups, *Comm. Alg.*, **18** (1990), 3423–3435.

FW_4 T.H. Fay and G.L. Walls. Compact nilpotent groups, *Comm. Alg.*, **17(9)** (1989), 2255–2268.

FW_5 T.H. Fay and G.L. Walls. Categorically compact locally nilpotent groups, a corrigendum, *Comm. Alg.*, **20** (1992), 1019–1022.

FW_6 T.H. Fay and G.L. Walls. A characterization of categorically compact locally nilpotent groups, preprint.

G_1 E. Giuli. Bases of topological epireflections, *Topology and its Appl.*, **27** (1980), 265–273.

G_2 E. Giuli. On m-separated projection spaces, *Journal of Applied Categorical Structures*, **2** (1994), 91–99.

G_3 E. Giuli. Compact pairs and generalized perfect maps, preprint.

GH_1 E. Giuli and M. Husek. A diagonal theorem for epireflective subcategories of **Top** and co-wellpoweredness, *Ann. Mat. Pura Appl.*, **IV**, **CXLV** (1986), 337–346.

GH_2 E. Giuli and M. Husek. A counterpart of compactness, *Bull. Un. Mat. Ital. B(7)*, **11(3)**, 605–621.

GMT E. Giuli, S. Mantovani and W. Tholen. Objects with closed diagonals, *J. Pure Appl. Algebra*, **51** (1988), 129–140.

GS E. Giuli and P. Simon. On spaces in which every bounded subset is Hausdorff, *Topology and its Appl.*, **37** (1990), 267–274.

GT E. Giuli and W. Tholen. Openness with respect to a closure operator, preprint.

H H. Herrlich. Topologische Reflexionen und Coreflexionen, *Springer L.N.M.*, **78**, Berlin, 1968.

Ho_1 D. Holgate. Closure operators in categories, Master Thesis, University of Cape Town, 1992.

Ho_2 D. Holgate. The pullback closure operator and generalization of perfectness, *Journal of Applied Categorical Structures*, **4** (1996), 107–120.

HP M. Hušek and D. Pumplün. Disconnectednesses, *Quaestiones Mathematicae*, **13** (1990), 449–459.

HS H. Herrlich and G.E. Strecker. *Category Theory*, 2nd ed. Berlin, Helderman Verlag, 1979.

HSS H. Herrlich, G. Salicrup and G.E. Strecker. Factorizations, denseness, separation, and relatively compact objects, *Topology and Its Appl.*, **27** (1987), 157–169.

K J. Koslowski. Closure operators with prescribed properties, *Category Theory and its Applications* (Louvain-la-Neuve, 1987) Springer L.N.M., **1248** (1988), 208–220.

L H. Lord. Functionally Hausdorff spaces, *Cahiers Topologie Geom. Differentielle Categoriques*, **30** (1989), 247–256.

NW L.D. Nel and R.G. Wilson. Epireflections in the category of T_0 spaces, *Fund. Math.*, **75** (1972), 69–74.

P D. Petz. Generalized connectedness and disconnectedness in topology, *Ann. Univ. Sci. Budapest Eötvös*, Sect. Math., **24** (1981), 247–252.

PR D. Pumplün and H. Röhrl. Separated totally convex spaces, *Manuscripta Math.*, **50** (1985), 145–183.

Pr_1 G. Preuß. Eine Galois-Korrespondenz in der Topologie, *Monatsh. Math.*, **75** (1971), 447–452.

Pr_2 G. Preuß. Relative connectednesses and disconnectednesses in topological categories, *Quaestiones Mathematicae*, **2** (1977), 297–306.

Pr_3 G. Preuß. Connection properties in topological categories and related topics, *Springer L.N.M.*, **719** (1979), 293–305.

S S. Salbany. Reflective subcategories and closure operators, *Proceedings of the Conference in Categorical Topology* (Mannheim, 1975), Springer L.N.M., **540** (1976), 548–565.

So_1 L. Sousa. α-sober spaces via the orthogonal closure operator, *Appl. Categorical Structures*, **4** (1996), 87–95.

So_1 L. Sousa. Orthogonality and closure operators, Cahiers Topologie Geom. Differentielle Categoriques, to appear.

St L. Stramaccia. Some remarks on closure operators induced by topological epireflections, preprint.

SV G. Salicrup and R. Vásquez. Connection and disconnection, *Springer L.N.M.*, **719** (1979), 326–344.

T_1 W. Tholen. Diagonal theorems in topology and elsewhere, *Proceedings Topological Symposium*, Prague, 1986, Heldermann, Berlin, 559–566.

T_2 W. Tholen. Factorizations, fibres and connectedness, *Proceedings of the Conference in Categorical Topology*, (Toledo, 1983), Helderman Verlag, Berlin, 1984, 549–566.

V H. V. Velichko, H-closed topological spaces, *Mat.Sb. (N.S.)*, **70** (1966), 98-i-112; *Amer. Mat. Soc. Transl.*, **78** (1969), 103–118.

G. Castellini
Department of Mathematics
University of Puerto Rico, Mayagüez Campus
P.O. Box 9018, Mayagüez, PR 00681-9018, U.S.A.
`castell@cs.uprm.edu`

Extensions of Maps from Dense Subspaces

H. L. Bentley

ABSTRACT In 1971, Douglas Harris introduced a concept that he called a $\mathcal{WO}$-map in order to find a subcategory of the category of $\mathbf{T}_1$ topological spaces for which the Wallman compactification becomes functorial. Later, Bentley and Naimpally generalized Harris' result to the setting of topological spaces endowed with the structure of a separating base in the sense of Steiner. In the present paper, the generalization is carried one step further to the setting of nearness spaces in the sense of Herrlich. Thus, it is shown that by restricting the class of maps, but not the class of spaces, the (strict) completion of a nearness space becomes functorial.

Key Words: Nearness space, completeness, completion, uniform continuity, superb map, excellent map, sleek map, $\mathcal{WO}$-map.

AMS Subject Classifications (2000): 54E17, 54D35, 54B30, 54D80.

"And things ain't like what they used to be."
– Blue Oyster Cult [1]

1 Conceptual Background

We are concerned not only with topological spaces, but also with a more general structure of a "topological" nature — namely with the type of structure that Herrlich called "nearness spaces". Also we are concerned more generally with the problem of extending a map defined on a dense subspace to a map defined on the whole space. A classical result of this nature is the theorem of Taĭmanov.

Theorem 1.1 (Taĭmanov's Theorem [Ta52]) *Let X be a topological space and let A be a dense subspace of X. Let Y be a compact Hausdorff space and let $f : A \to Y$ be a continuous map. Then f has a continuous extension*

[1] From *Fortune* April 9, 1996, p. 181.

$g : X \to Y$ if and only if whenever $\mathcal{B}$ is a finite collection of closed subsets of Y with $\cap\mathcal{B} = \emptyset$, then we also have

$$\bigcap_{B \in \mathcal{B}} \mathrm{cl}_X f^{-1}B = \emptyset.$$

The Taǐmanov Theorem is a result *par excellence* of this type within the setting of general topological spaces with a compact Hausdorff range — one can hardly expect to improve upon it. Nevertheless, by introducing richer kinds of structures one gains a fuller understanding as to why the conditions involved in the Taǐmanov Theorem not only work but are natural.

In 1974, Herrlich proved a generalization of Taǐmanov's Theorem in the setting of nearness spaces for maps with a regular codomain.

Theorem 1.2 (Herrlich's Theorem [He74b]) *Let X be a nearness space and let A be a dense subset of X. Let Y be a complete, regular nearness space and let $f : A \to Y$ be a map. Then f can be uniformly continuously extended to X iff whenever a collection $\mathcal{B}$ is far in Y then $f^{-1}\mathcal{B} = \{ f^{-1}B \mid B \in \mathcal{B} \}$ is far in X.*

A corollary to Herrlich's Theorem is the following very neat generalization of the Taǐmanov Theorem.

Corollary 1.3 (Herrlich [He74b]) *Let X be a symmetric topological space and let A be a dense subspace of X. Let Y be a regular topological space and let $f : A \to Y$ be a continuous map. Then f has a continuous extension $g : X \to Y$ if and only if whenever $\mathcal{B}$ is a collection of closed subsets of Y with $\cap\mathcal{B} = \emptyset$, then we also have*

$$\bigcap_{B \in \mathcal{B}} \mathrm{cl}_X f^{-1}B = \emptyset.$$

While the above theorems are very nice results, they suffer from the defect that the codomain of the map to be extended is required to be a regular space. For the non-regular codomain case, only a few results are known:

(1) A closed surjection between $\mathbf{T}_1$ spaces can be extended to a closed surjection between the Wallman compactifications (Arhangel'skii) (reported by Ponomarev [Po64]).

(2) A $\mathcal{WO}$-map between $\mathbf{T}_1$ spaces can be extended to a $\mathcal{WO}$-map between the Wallman compactifications (Harris [Ha71]).

(3) A $\mathcal{WO}$-map between $\mathbf{T}_1$ spaces endowed with a separating base a la Steiner [St68] can be extended to a $\mathcal{WO}$-map between the Wallman-type compactifications (Bentley and Naimpally [BN74a]).

(4) The Wallman–Shanin–type compactification of Lowen's approach spaces can be viewed as an epireflection (Sioen [Si00]).

Each of the four results mentioned above represents a generalization of each of the preceding results.

In this paper, we shall present a further generalization of the first three of these results in the setting of nearness spaces.

2 Technical Background

Herein are gathered the various details we need concerning nearness spaces and their completions. Nearness spaces were introduced by Herrlich [He74a]. In [He74c] Herrlich showed how his nearness structures simultaneously generalize both topological spaces and uniform spaces, and how they permit a meaningful concept of cauchyness for filters (and the more general "micromeric" collections of Katětov [Ka63, Ka65]). For clarity, we present the definition of nearness spaces here, but for a full description of these structures, reference is made either to the papers on nearness spaces listed in the extensive bibliography which appears in Herrlich's survey [He83], or to his textbook [He88].

A nearness space can have its structure described in at least four ways, using as a primitive one of:

(a) The collections which are near.

(b) The collections which are far.

(c) The collections which are micromeric (i.e., cauchy).

(d) The uniform covers.

The approach used by Herrlich [He74a] in his original paper on nearness involved mainly the first two above concepts. For details regarding the relationships between the four concepts (near, far, micromeric, uniform cover) we refer the reader to [He74c].

Definition 2.1 A *nearness space* consists of a set X together with a distinguished set of collections of subsets of X called *near collections* subject to the following axioms:

(1) If a collection $\mathcal{A}$ is such that $\cap\mathcal{A} \neq \emptyset$, then $\mathcal{A}$ is near.

(2) If $\emptyset \in \mathcal{A}$, then $\mathcal{A}$ is not near (collections which are not near are referred to as being *far*).

(3) If we have collections $\mathcal{A}$ and $\mathcal{B}$ such that $\mathcal{A}$ corefines [2] $\mathcal{B}$, then $\mathcal{A}$ is near if $\mathcal{B}$ is.

(4) If we have collections $\mathcal{A}$ and $\mathcal{B}$ such that $\mathcal{A} \vee \mathcal{B} = \{ A \cup B \mid A \in \mathcal{A}$ and $B \in \mathcal{B}\}$ is near, then $\mathcal{A}$ is near or $\mathcal{B}$ is near.

(5) If we define the closure operator cl on X by cl $A = \{x \in X \mid \{\{x\}, A\}$ is near $\}$ and we define for a collection $\mathcal{A}$, cl $\mathcal{A} = \{\text{cl } A \mid A \in \mathcal{A}\}$, then $\mathcal{A}$ is near if cl $\mathcal{A}$ is.

A map $f : X \to Y$ between nearness spaces X and Y is said to be *uniformly continuous* iff whenever a collection $\mathcal{A}$ is near in X, then the collection of all images $f\mathcal{A} = \{ fA \mid A \in \mathcal{A} \}$ is near in Y.

We denote by *Near* the resulting concrete category of all nearness spaces and uniformly continuous maps. It is a *topological category* in the sense of Herrlich [He74d].

It turns out that the closure operator defined in (5) associates with each nearness space an underlying topological structure. The underlying topological structure of a nearness space always satisfies the symmetry[3] separation axiom, an axiom slightly weaker than the $\mathbf{T}_1$ axiom.

If we denote by $\mathbf{Top_S}$ the category of all symmetric topological spaces, then $\mathbf{Top_S}$ is embeddable into **Near** as a bicoreflective full subcategory by the functor which assigns to each symmetric topological space X the nearness space consisting of X, with the nearness structure defined by the prescription that a collection $\mathcal{A}$ of subsets of X is *near* iff, for the collection cl $\mathcal{A} = \{ \text{cl}_X A \mid A \in \mathcal{A} \}$ of all closures in X of members of $\mathcal{A}$, we have that $\bigcap \text{cl}\mathcal{A} \neq \emptyset$. By means of this embedding, we assume that $\mathbf{Top_S}$ is actually a subcategory of **Near**, i.e., that every symmetric topological space is a nearness space with nearness structure defined as above.

Furthermore, if we denote by **Unif** the category of all uniform spaces, then **Unif** is embeddable into **Near** as a bireflective full subcategory by the functor which assigns to each uniform space X the nearness space consisting of X, with the nearness structure defined by the prescription that a collection $\mathcal{A}$ of subsets of X is *near* iff, for each uniform cover $\mathcal{U}$ of X, there exists a member U of $\mathcal{U}$ which meets every member of $\mathcal{A}$.

Finally, if we denote by **Mer** the category of all merotopic spaces (Katětov [Ka63], [Ka65]), then **Near** is embeddable into **Mer** as a bireflective full subcategory by the functor which assigns to each nearness space X, the merotopic space having the same underlying set, and with the mero-

[2] We say that $\mathcal{A}$ *corefines* $\mathcal{B}$ iff every member of $\mathcal{A}$ contains as a subset some member of $\mathcal{B}$.

[3] A *symmetric* topological space is one which satisfies $x \in \text{cl}\{y\} \Longleftrightarrow y \in \text{cl}\{x\}$. These spaces were first considered by Šanin [Ša43] who used the term *weakly regular.* They have been called R_0-*spaces* by A. S. Davis [Da61].

topic structure being defined by the prescription that a collection $\mathcal{A}$ of subsets of X is *micromeric* iff the collection

$$\sec \mathcal{A} = \{\ B \subset X \mid A \cap B \neq \emptyset \text{ for all } A \in \mathcal{A}\ \}$$

is near. A micromeric filter is called a *cauchy*[4] filter. It follows easily that this notion of cauchyness for a filter coincides with the usual notion if X is a uniform space. Intuitively, the notion of being micromeric can be thought of as containing *arbitrarily small sets.*

We need to say what it means for a nearness space to be complete and regular.

Definition 2.2 A nearness space is said to be *regular* if it satisfies the following condition: Whenever a collection $\mathcal{A}$ is far, then so is the collection

$$\{\ B \subset X \mid A < B \text{ for some } A \in \mathcal{A}\ \}$$

where by $A < B$ we mean that $\{A, X \setminus B\}$ is far.

Conveniently, but of course not coincidentally, it turns out that a topological space is regular iff it is regular in the usual topological sense. The category of regular nearness spaces coincides with Morita's semi-uniform spaces [Mo51]. They were developed further by Steiner and Steiner [SS73]. See also [BH96a]. In [BH79a] regular nearness spaces, which are more general than uniform spaces, were shown to be as pleasant as uniform spaces in many important ways.

Definition 2.3 A *cluster* in a nearness space X is a nonempty maximal near collection. Among the clusters are those of the form

$$e_X(x) = \{\ B \subset X \mid x \in \mathrm{cl} B\ \}$$

for $x \in X$; clusters of this form are called *fixed clusters.* There may be clusters which are not fixed; such ones are called *free clusters.* If no free clusters exist in X then X is said to be *complete.*

For a regular nearness space (a fortiori, also for a uniform space) the above concept of completeness coincides with the usual one that every cauchy filter converges [He74a]. We let **Compl** denote the full subcategory of **Near** consisting of the complete spaces.

It turns out that every symmetric topological space is a complete nearness space.

A nearness space is said to be a $\mathbf{T}_1$ space if its underlying topological space is $\mathbf{T}_1$ in the usual sense, i.e., if $\{\{x\},\{y\}\}$ is near implies $x = y$.

[4]Perhaps a supreme honor for a mathematician is that a concept be labeled with his uncapitalized name, because the concept is so broadly used.

We let $\mathbf{Near}_1$ denote the full subcategory of $\mathbf{Near}$ consisting of the $\mathbf{T}_1$ nearness spaces, and similarly, we let $\mathbf{Compl}_1$ denote the full subcategory of $\mathbf{Near}$ consisting of the complete $\mathbf{T}_1$ nearness spaces.

In [He74a] Herrlich showed that every nearness space has a completion. We need the details of his construction.

The completion of a nearness space X is defined by first letting X^* denote the set of all clusters on X and then, for a collection Ω of subsets of X^*, we define Ω to be *near* in X^* provided the collection

$$\cup\{\cap\omega \mid \omega \in \Omega\}$$

is near in X. With this structure, X^* becomes a nearness space. The map $e_X : X \to X^*$ defined earlier by

$$e_X(x) = \{ B \subset X \mid x \in \mathrm{cl}B \} \quad \text{for } x \in X,$$

is uniformly continuous and its image is dense in X^*.

The following hold:

(1) $\mathcal{G} \in \mathrm{cl}_{X^*}\omega \iff \cap\omega \subset \mathcal{G}$.

(2) $\mathcal{G} \in \mathrm{cl}_{X^*}e_X[G] \iff G \in \mathcal{G}$.

(3) $A \in \cap\omega \iff \omega \subset \mathrm{cl}(e_X A)$.

(4) $B \in \cap(e_X A) \iff A \subset \mathrm{cl}B$.

for all $\mathcal{G} \in X^*$, $\omega \subset X^*$, and $A, B, G \subset X$. (For the proofs of these assertions as well as for the proof of the following proposition, see [He74a]).

Proposition 2.4 *Let X be a nearness space. Then:*

(1) X^* is a complete $\mathbf{T}_1$ space.

(2) $\mathcal{A}$ is near in X iff $e_X\mathcal{A} = \{ e_X A \mid A \in \mathcal{A} \}$ is near in X^*.

(3) If X is $\mathbf{T}_1$ then $e_X : X \to X^*$ is a uniform embedding.

3 Clumps

Herrlich proved that the category of complete, regular $\mathbf{T}_1$ nearness spaces is reflective in $\mathbf{Near}$, and is epireflective in $\mathbf{Near}_1$. Our objective is to get as close to these results as we can without the restriction of regularity.

Clusters are even more elusive than ultrafilters — Zorn's Lemma usually is of no help in trying to find a cluster containing a given near collection. One remedy for this difficulty is the concept of a clump. A clump automatically determines a cluster which contains it.

Definition 3.1

(1) A collection of subsets of a nearness space is said to be *concentrated* if it is both near and micromeric.

(2) A *clump* in a nearness space X is a concentrated collection $\mathcal{M}$ such that the collection

$$\text{cluster}(\mathcal{M}) = \{\ E \subset X \mid \{E\} \cup \mathcal{M} \text{ is near}\}$$

is near.

Proposition 3.2 *If $\mathcal{A}$ is near and cluster $(\mathcal{A})$ is nonempty and near, then cluster $(\mathcal{A})$ is a cluster.*

Proof. Since $\mathcal{A}$ is near, we have

$$A \in \mathcal{A} \quad \Longrightarrow \quad \{A\} \cup \mathcal{A} = \mathcal{A} \text{ is near} \quad \Longrightarrow \quad A \in \text{cluster}(\mathcal{A}).$$

Therefore $\mathcal{A} \subset \text{cluster}(\mathcal{A})$. Next, if $\{B\} \cup \text{cluster}(\mathcal{A})$ is near, then the fact that $\{B\} \cup \mathcal{A} \subset \{B\} \cup \text{cluster}(\mathcal{A})$ implies that $\{B\} \cup \mathcal{A}$ is near, which in turn implies that $B \in \text{cluster}(\mathcal{A})$. So, cluster $(\mathcal{A})$ is a maximal near collection. ■

Proposition 3.3 *Every cluster is a clump.*

Proposition 3.4 *Let X be a nearness space and let $\mathcal{M}$ be a clump in X. Then:*

(1) $\mathcal{M} \subset \text{cluster}(\mathcal{M})$.

(2) cluster $(\mathcal{M})$ is a cluster in X.

(3) $\mathcal{M}$ is a cluster iff $\mathcal{M} = \text{cluster}(\mathcal{M})$.

(4) If $\mathcal{A}$ is near in X with $\mathcal{M} \subset \mathcal{A}$ then we have that for all $E \subset X$, $\{E\} \cup \mathcal{M}$ is near iff $\{E\} \cup \mathcal{A}$ is.

(5) If $\mathcal{G}$ is a cluster in X with $\mathcal{M} \subset \mathcal{G}$ then we have that cluster $(\mathcal{M}) = \mathcal{G}$.

(6) If $\mathcal{A}$ is near and $\mathcal{M} \subset \mathcal{A}$ then $\mathcal{A}$ is a clump in X.

Definition 3.5 If $\mathcal{M}$ is a clump, then cluster $\mathcal{M}$ is called the *cluster generated by $\mathcal{M}$.*

Remarks 3.6

(1) In a regular space every collection which is concentrated is a clump.

(2) Even in a regular space, it can very well happen that for a near collection $\mathcal{M}$, the collection $\{ E \subset X \mid \{E\} \cup \mathcal{M} \text{ is near } \}$ is near but $\mathcal{M}$ fails to be micromeric. For example, let X be the set of all non-negative integers with the discrete nearness structure (i.e., $\mathcal{A}$ is near iff $\cap\mathcal{A} \neq \emptyset$) and let

$$\mathcal{M} = \{ \{0\} \cup \{ m \in X \mid m \geq n \} \mid n \in X \}.$$

Then $\mathcal{M}$ is not micromeric since the collection $\mathcal{B} = \{ \{0\}, X \setminus \{0\} \}$ is far but every element of $\mathcal{M}$ meets every element of $\mathcal{B}$.

(3) In a discrete space X, for a near collection $\mathcal{M}$ the following are equivalent:

(a) $\mathcal{M}$ is a clump.

(b) $\mathcal{M}$ is micromeric.

(c) For some $x \in X$ we have $\{x\} = \cap\mathcal{M} = \cap \sec \mathcal{M}$.

(d) For some $x \in X$ we have $\{x\} \in \mathcal{M}$.

Proposition 3.7 *Let X be a nearness space and X^* its completion. Then:*

(1) If Ω is a cluster in X^*, then $\cup\{\cap\omega \mid \omega \in \Omega \}$ is a cluster in X.

(2) If Λ is concentrated in X^*, then $\cup\{\cap\lambda \mid \lambda \in \Lambda \}$ is concentrated in X.

(3) If Λ is a clump in X^*, then $\cup\{\cap\lambda \mid \lambda \in \Lambda \}$ is a clump in X.

Proof.

(1): Let $\mathcal{G} = \cup\{\cap\omega \mid \omega \in \Omega \}$. Since Ω is near, so is $\mathcal{G}$. To show $\mathcal{G}$ is maximal, assume that $\{G\}\cup\mathcal{G}$ is near. Since $\cap e_X[G] = \{ E \subset X \mid G \subset \mathrm{cl}E \}$ it follows that cl $[(\cap e_X[G]) \cup \mathcal{G}]$ corefines $\{G\} \cup \mathcal{G}$. Therefore $(\cap e_X[G]) \cup \mathcal{G}$ is near, and from that, by definition of the structure of X^*, we get that $\{e_X[G]\} \cup \Omega$ is near. Thus $e_X[G] \in \Omega$ and since $G \in \cap e_X[G]$ we have $G \in \mathcal{G}$ as required.

(2): Let $\mathcal{M} = \cup\{\cap\lambda \mid \lambda \in \Lambda \}$. $\mathcal{M}$ is near. We must show that $\mathcal{M}$ is micromeric. That will follow immediately if we can show that

$$\sec \mathcal{M} \subset \cup\{\cap\tau \mid \tau \in \sec \Lambda\}.$$

Toward that end, let $B \in \sec \mathcal{M}$. Then $X \backslash B \neq \mathrm{stack}\mathcal{M}$ so for all $\lambda \in \Lambda$, $X\backslash B \notin \cap\lambda$; choose $\mathcal{G}_\lambda \in \lambda$ with $X\backslash B \notin \mathcal{G}_\lambda$ and let $\tau = \{ \mathcal{G}_\lambda \mid \lambda \in \Lambda \}$. Then $B \in \cap\tau$ and $\tau \in \sec \Lambda$.

(3): Let $\mathcal{M} = \cup\{\cap\lambda \mid \lambda \in \Lambda \}$. By (2), $\mathcal{M}$ is concentrated. Let $\Omega = \mathrm{cluster}(\Lambda)$ and $\mathcal{G} = \mathrm{cluster}(\mathcal{M})$. Then Ω is near, and if we can show that $\mathcal{G} \subset \cup\{\cap\omega \mid \omega \in \Omega \}$ it will follow that $\mathcal{G}$ is near, as required. Toward that end, let $G \in \mathcal{G}$. Then $\{G\} \cup \mathcal{M}$ is near. As in the proof in (1) above, we have that cl $[(\cap e_X[G]) \cup \mathcal{M}]$ corefines $\{G\} \cup \mathcal{M}$. Therefore $\{e_X[G]\} \cup \Lambda$ is near, from which we get that $e_X[G] \in \Omega$. Since $G \in \cap e_X[G]$ we have $G \in \cup\{\cap\omega \mid \omega \in \Omega\}$ as required. ■

4 Supurb, Excellent and Sleek Maps

Since not every uniformly continuous map $f : X \to Y$ has a uniformly continuous extension $X^* \to Y^*$ to the completions, we shall be studying certain non-full subcategories of **Near**. The definition of certain restricted classes of maps is preceded by a preliminary concept.

Definition 4.1 Let $\mathcal{D}$ and $\mathcal{G}$ be collections of subsets of a nearness space X. We say that $\mathcal{D}$ and $\mathcal{G}$ *embrace* provided that for all $E \subset X$,

$$\{E\} \cup \mathcal{D} \textit{ is near} \qquad \Longleftrightarrow \qquad \{E\} \cup \mathcal{G} \textit{ is near.}$$

Example 4.2

(1) In a discrete space, two collections $\mathcal{D}$ and $\mathcal{G}$ embrace iff $\cap\mathcal{D} = \cap\mathcal{G}$.

(2) In an indiscrete space, two collections $\mathcal{D}$ and $\mathcal{G}$ embrace iff both $\mathcal{D}$ and $\mathcal{G}$ are near or both $\mathcal{D}$ and $\mathcal{G}$ are far.

Remarks 4.3 Using the "embracing" terminology, Proposition 3.4 yields the following statements:

(1) If $\mathcal{M}$ is a clump, $\mathcal{A}$ is near, and $\mathcal{M} \subset \mathcal{A}$, then $\mathcal{M}$ and $\mathcal{A}$ embrace.

(2) If $\mathcal{M}$ is a clump and $\mathcal{G}$ is a cluster, then $\mathcal{M}$ and $\mathcal{G}$ embrace iff $\mathcal{M} \subset \mathcal{G}$.

(3) If $\mathcal{M}$ is a clump and $\mathcal{G} = \text{cluster}(\mathcal{M})$, then $\mathcal{M}$ and $\mathcal{G}$ embrace.

Definition 4.4 Let $f : X \to Y$ be a map between nearness spaces. Then we say that:

(1) $f : X \to Y$ is a *superb* map iff whenever $\mathcal{B}$ is far in Y there exists $\mathcal{A}$ far in X and there exists a function

$$\mathcal{A} \to \mathcal{B}, \qquad A \mapsto B_A$$

such that for every clump $\mathcal{M}$ in X and for every $A \in \mathcal{A}$,

$$\{A\} \cup \mathcal{M} \text{ is far} \qquad \Longrightarrow \qquad \{B_A\} \cup f\mathcal{M} \text{ is far.}$$

(To simplify the language, we shall say that the function $\mathcal{A} \to \mathcal{B}$, $A \mapsto B_A$ *demonstrates the superbness* of f iff for every clump $\mathcal{M}$ in X and for every $A \in \mathcal{A}$,

$$\{A\} \cup \mathcal{M} \text{ is far} \qquad \Longrightarrow \qquad \{B_A\} \cup f\mathcal{M} \text{ is far}$$

holds.)

(2) $f : X \to Y$ is an *excellent* map iff f is uniformly continuous and whenever $\mathcal{M}$ is a clump in X, then $f\mathcal{M}$ is a clump in Y.

(3) $f : X \to Y$ is a *sleek* map iff f is uniformly continuous and whenever $\mathcal{M}$ and $\mathcal{N}$ are clumps in X which embrace, then $f\mathcal{M}$ and $f\mathcal{N}$ embrace in Y.

Remarks 4.5 It would be interesting to investigate the following condition, which is stronger than sleekness: Whenever $\mathcal{M}$ and $\mathcal{N}$ are near collections in X which embrace, then $f\mathcal{M}$ and $f\mathcal{N}$ embrace in Y. In a way, this concept is more natural than is uniform continuity itself. However, we leave the project of investigating that stronger concept for another time — for the purposes of this paper, we needed only the weaker formulation (involving clumps) that we used to define sleekness.

Proposition 4.6

(1) Every superb map is excellent.

(2) Every excellent map is sleek.

Proof.

(1): Let $f : X \to Y$ be a superb map. First we show that f is uniformly continuous. Let $\mathcal{B}$ be far in Y. There exists a far collection $\mathcal{A}$ in X and a function $\mathcal{A} \to \mathcal{B}$, $A \mapsto B_A$ demonstrating the superbness of f. To establish that $f^{-1}\mathcal{B}$ is far in X, it will be sufficient to show that for all $A \in \mathcal{A}$, we have $f^{-1}[B_A] \subset \mathrm{cl}A$. To that end, let $x \in f^{-1}[B_A]$. Then $\mathcal{M} = \{\{x\}\}$ is a clump in X. Since $\{B_A, \{f(x)\}\}$ is near it follows that $\{B_A\} \cup f\mathcal{M}$ is near. Therefore, $\{A\} \cup \mathcal{M}$ is near and so $x \in \mathrm{cl}A$ as required.

Next we show that f is excellent. Let $\mathcal{M}$ be a clump in X. Since f is uniformly continuous, we have that $f\mathcal{M}$ is concentrated. Let $\mathcal{B} = \mathrm{cluster}(f\mathcal{M})$ and suppose that $\mathcal{B}$ is far in Y. Then there exists $\mathcal{A}$ far in X and a function $\mathcal{A} \to \mathcal{B}$, $A \mapsto B_A$ which demonstrates the superbness of f. Since $\mathcal{A}$ is far and cluster $(\mathcal{M})$ is near, there exists $A \in \mathcal{A}$ with $A \notin \mathrm{cluster}(\mathcal{M})$. Thus $\{A\} \cup \mathcal{M}$ is far in X. Therefore $\{B_A\} \cup f\mathcal{M}$ is far in Y, which contradicts the fact that $B_A \in \mathcal{B}$.

(2): Let $f : X \to Y$ be an excellent map. Let $\mathcal{M}$ and $\mathcal{N}$ be clumps in X which embrace. Let $\mathcal{G} = \mathrm{cluster}(\mathcal{M}) = \mathrm{cluster}(\mathcal{N})$. f is excellent so $f\mathcal{M}$ and $f\mathcal{N}$ are clumps. $\mathcal{G}$, being a cluster, is also a clump. Hence $f\mathcal{G}$ is a clump too. Since $f\mathcal{M} \subset f\mathcal{G}$ it follows that

$$\mathrm{cluster}(f\mathcal{G}) \subset \mathrm{cluster}(f\mathcal{M}).$$

Since both are clusters, we have the equality

$$\mathrm{cluster}(f\mathcal{G}) = \mathrm{cluster}(f\mathcal{M}).$$

Similarly,

$$\mathrm{cluster}(f\mathcal{G}) = \mathrm{cluster}(f\mathcal{N}),$$

and therefore we have cluster$(f\mathcal{M})$ = cluster$(f\mathcal{N})$. From this equation, the sleekness of f follows immediately. ■

Proposition 4.7 *Let X and Y be any nearness spaces and let $f : X \to Y$ be a map. Then for the following three statements we have that (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3).*

(1) f is excellent.

(2) If $\mathcal{M}$ and $\mathcal{N}$ are two clumps in X which generate the same cluster, then $f\mathcal{M}$ and $f\mathcal{N}$ are clumps in Y which generate the same cluster.

(3) $f : X \to Y$ is superb iff the following condition holds: whenever $\mathcal{B}$ is far in Y, there exists $\mathcal{A}$ far in X and there exists a function

$$\mathcal{A} \to \mathcal{B}, \qquad A \mapsto B_A$$

such that for every cluster $\mathcal{G}$ in X and for every $A \in \mathcal{A}$,

$$\{A\} \cup \mathcal{G} \text{ is far} \qquad \Longrightarrow \qquad \{B_A\} \cup f\mathcal{G} \text{ is far.}$$

(Remark: Note that the condition in (3) is identical with the definition of superbness except for the replacement of the term "clump" by the term "cluster".)

Proof. (1)$\Longrightarrow$(2): Let $\mathcal{M}$ and $\mathcal{N}$ be two clumps in X that generate the same cluster; denote it by $\mathcal{G}$ = cluster$(\mathcal{M})$ = cluster$(\mathcal{N})$. We have that for all $A \subset X$,

$$\{A\} \cup \mathcal{M} \text{ is near} \qquad \Longleftrightarrow \qquad A \in \mathcal{G} \qquad \Longleftrightarrow \qquad \{A\} \cup \mathcal{N} \text{ is near}$$

from which it follows that $\mathcal{M}$ and $\mathcal{N}$ embrace. Since f is sleek, $f\mathcal{M}$ and $f\mathcal{N}$ embrace. Since f is excellent, we have that $f\mathcal{M}$ and $f\mathcal{N}$ are clumps in Y. Since these two clumps embrace, they generate the same cluster.

(2)$\Longrightarrow$(3): Clearly if f is superb, then the stated condition is satisfied. For the proof in the reverse direction, assume that the stated condition is satisfied. Let $\mathcal{B}$ be far in Y and let $\mathcal{A}$ be the far collection in X guaranteed to exist by the condition in (3), with the function $\mathcal{A} \to \mathcal{B}$, $A \mapsto B_A$ being one which satisfies the stated condition with respect to clusters. We shall show that it demonstrates the superbness of f. To that end, let $\mathcal{M}$ be a clump in X and let $A \in \mathcal{A}$ such that $\{A\} \cup \mathcal{M}$ is far. Let $\mathcal{G}$ = cluster$(\mathcal{M})$. Then $\mathcal{M}$ and $\mathcal{G}$ are clumps which generate the same cluster, namely $\mathcal{G}$. By condition (2), we have that $f\mathcal{M}$ and $f\mathcal{G}$ generate the same cluster in Y, call it $\mathcal{H}$. $A \notin \mathcal{G}$ so $\{A\} \cup \mathcal{G}$ is far. Since $\mathcal{G}$ is a cluster, $\{B_A\} \cup f\mathcal{G}$ is far in Y. Thus $B_A \notin \mathcal{H}$ and therefore $\{B_A\} \cup f\mathcal{M}$ is far. ■

Proposition 4.8 *Let X and Y be nearness spaces with Y regular. For a map $f : X \to Y$, the following are equivalent:*

(1) f is superb.

(2) f is excellent.

(3) f is sleek.

(4) f is uniformly continuous.

Proof. By Proposition 4.6 we have (1)$\Longrightarrow$(2) $\Longrightarrow$(3). By definition, (3)$\Longrightarrow$(4). We shall prove that (4)$\Longrightarrow$(1). Assume that $f : X \to Y$ is uniformly continuous. Let $\mathcal{B}$ be far in Y and let

$$\mathcal{D} = \{ \, D \subset Y \mid B < D \text{ for some } B \in \mathcal{B} \, \}.$$

Since Y is regular, $\mathcal{D}$ is far and hence $\mathcal{A} = f^{-1}\mathcal{D}$ is far in X. For each $A \in \mathcal{A}$ choose a member $D_A \in \mathcal{D}$ with $A = f^{-1}[D_A]$ and then choose a member $B_A \in \mathcal{B}$ with $B_A < D_A$. We claim that the function $A \mapsto B_A$, $\mathcal{A} \to \mathcal{B}$ demonstrates the superbness of f. To show this, let $A \in \mathcal{A}$ and let $\mathcal{M}$ be a clump in X with $\{A\} \cup \mathcal{M}$ far in X. Since $B_A < D_A \subset Y \setminus f[X \setminus A]$, then $\{B_A, f[X \setminus A]\}$ is far. Since $\mathcal{M}$ is micromeric and f is uniformly continuous, it follows that $f\mathcal{M}$ is micromeric. Since Y is regular, $f\mathcal{M}$ is micromeric, and $\{B_A, f[X \setminus A]\}$ is far; there exists an element $M \in \mathcal{M}$ such that either $\{fM, B_A\}$ is far or $\{fM, f[X \setminus A]\}$ is far. The latter cannot happen since $\{X \setminus A\} \cup \mathcal{M}$ is near which implies that $\{f[X \setminus A]\} \cup f\mathcal{M}$ is near, and $\{fM, f[X \setminus A]\}$ being a subcollection of that, is also near. Therefore, we have that $\{fM, B_A\}$ is far and, since that collection is a subcollection of $\{B_A\} \cup f\mathcal{M}$, the latter must also be far. Hence f is superb. ∎

Lemma 4.9 *Let X and Y be nearness spaces and $f : X \to Y$ a map. Let $x \in X$. Then we have:*

(1) cluster $(f[e_X(x)]) \subset e_Y(f(x))$.

(2) If f is uniformly continuous, then cluster $(f[e_X(x)]) = e_Y(f(x))$.

Proof.

(1): Let $B \in \text{cluster}(f[e_X(x)])$. Then $\{B\} \cup f[e_X(x)]$ is near. Since $\{x\} \in e_X(x)$, we have $\{\{f(x)\}\} \subset f[e_X(x)]$. Consequently, $\{B, \{f(x)\}\} \subset \{B\} \cup f[e_X(x)]$. Therefore $\{B, \{f(x)\}\}$ is near. This means that $f(x) \in \text{cl}_Y B$ and consequently $B \in e_Y(f(x))$, as required.

(2): Since $e_X(x)$ is near in X and f is uniformly continuous, $f[e_X(x)]$ is near in Y. From $\{\{f(x)\}\} \subset f[e_X(x)]$ it follows that $\{\{f(x)\}\} \cup f[e_X(x)] = f[e_X(x)]$ is near in Y. Therefore we have $\{f(x)\} \in \text{cluster}(f[e_X(x)])$ and so the collection cluster $(f[e_X(x)])$ is nonempty. By Proposition 3.2 cluster$(f[e_X(x)])$ is a cluster. Since $e_Y(f(x))$ is a cluster, from (1) we get the desired equality. ∎

Proposition 4.10 *Let X and Y be nearness spaces with X complete. For a map $f : X \to Y$, the following are equivalent:*

(1) f is superb.

(2) f is excellent.

(3) f is sleek.

Proof. We have from Proposition 4.6 that (1)$\Longrightarrow$(2)$\Longrightarrow$(3). We show that (3)$\Longrightarrow$(2)$\Longrightarrow$(1).

(3)$\Longrightarrow$(2): Let $\mathcal{M}$ be a clump in X. Since X is complete, there exists $x \in X$ with $e_X(x) = \text{cluster}(\mathcal{M})$. By Proposition 3.4 (4) (see also Remarks 4.3 (3)) $\mathcal{M}$ and $e_X(x)$ embrace. f being excellent implies that $f\mathcal{M}$ and $f[e_X(x)]$ embrace. Therefore we have, for any subset B of X,

$$B \in \text{ cluster } (f\mathcal{M} \Longrightarrow \{B\} \cup f\mathcal{M} \text{ is near } \Longrightarrow \{B\} \cup f[e_X(x)] \text{ is near } \Longrightarrow B \in e_Y(f(x))$$

where the last step follows from (2) of the Lemma. Therefore, we have that cluster $(f\mathcal{M}) \subset e_Y(f(x))$. From Proposition 3.2 it follows that the collection cluster $(f\mathcal{M})$ is a cluster. Hence, $f\mathcal{M}$ is a clump, as was to be proved.

(2)$\Longrightarrow$(1): Assume that f is excellent. Let $\mathcal{B}$ be far in Y. Then cl $\mathcal{B}$ is far in Y and since f is uniformly continuous,

$$\mathcal{A} = \{\ f^{-1}[\text{cl}B] \mid B \in \mathcal{B}\ \}$$

is far in X. For each $A \in \mathcal{A}$ choose a $B_A \in \mathcal{B}$ such that $A = f^{-1}[\text{cl}B_A]$. This defines a function $\mathcal{A} \to \mathcal{B}$, $A \mapsto B_A$, which we claim demonstrates the superbness of f. Since f is excellent, it suffices to show that it satisfies the "cluster" condition in Proposition 4.7 (3). To that end, let $A \in \mathcal{A}$ and let $\mathcal{G}$ be a cluster in X such that $\{A\} \cup \mathcal{G}$ is far. Since X is complete, there exists $x \in X$ with $\{x\} \in \mathcal{G}$. From $A \notin \mathcal{G}$ it follows that $x \notin \text{cl}A$. Since f, being uniformly continuous, must be continuous with respect to the underlying topologies, we have that

$$x \notin \text{cl}A = \text{cl}f^{-1}[\text{cl}B_A] = f^{-1}[\text{cl}B_A].$$

Therefore, $\{B_A\} \cup \{\{f(x)\}\}$ is far and since $\{f(x)\} \in f\mathcal{G}$ then $\{B_A\} \cup f\mathcal{G}$ is far. Hence f is superb. ∎

We close this section with a few results relating the preceding kinds of maps with certain others that have appeared previously in the literature.

Definition 4.11 Let $f : X \to Y$ be a uniformly continuous map between nearness spaces. Then we say that:

(1) $f : X \to Y$ is an *initial* [5] map iff for any collection $\mathcal{A}$ of subsets of X, if $f\mathcal{A}$ is near in Y, then $\mathcal{A}$ was already near in X.

(2) $f : X \to Y$ is a *strict* [6] map iff each far collection $\mathcal{B}$ in Y is corefined by some collection of the form $\mathrm{cl}_Y f\mathcal{A}$ for some far collection $\mathcal{A}$ in X.

(3) $f : X \to Y$ is an *exclusive*[7] map iff whenever $\mathcal{G}$ is a near grill[8] on X and $B \subset Y$ such that $\{f^{-1}B\} \cup \mathcal{G}$ is far in X, then $\{B\} \cup f\mathcal{G}$ is far in Y.

Proposition 4.12

(1) Every surjective initial map is strict.

(2) Every strict initial map with nonempty domain is a dense map.

(3) Every strict initial map is superb.

Proof.
(1): Let $f : X \to Y$ be a surjective initial map and let $\mathcal{B}$ be far in Y. Define $\mathcal{A} = f^{-1}\mathcal{B}$, observe that $\mathcal{A}$ is far in X, and since f is surjective, $f\mathcal{A}$ equals, hence corefines, $\mathcal{B}$.

(2): Let $f : X \to Y$ be a strict initial map with nonempty domain. Let $y \in Y$ and suppose that $y \notin \mathrm{cl}_Y fX$. $\mathcal{B} = \{\{y\}, fX\}$ is far in Y so for some $\mathcal{A}$ far in X we have $\mathrm{cl}_Y f\mathcal{A}$ corefining $\mathcal{B}$. Since f is initial, $\mathrm{cl}_Y f\mathcal{A}$ is far in Y and therefore, since X is nonempty, we must have $fX \not\subset \mathrm{cl}_Y fA$ for some $A \in \mathcal{A}$. But then $y \in \mathrm{cl}_Y fA \subset \mathrm{cl}_Y fX$, a contradiction.

(3): Let $f : X \to Y$ be a strict initial map and let $\mathcal{B}$ be far in Y. Since f is strict, there exists $\mathcal{A}$ far in X such that $\mathrm{cl}_Y f\mathcal{A}$ corefines $\mathcal{B}$. For every $A \in \mathcal{A}$ select $B_A \in \mathcal{B}$ such that $B_A \subset \mathrm{cl}_Y fA$. We shall show that the function $A \mapsto B_A$, $\mathcal{A} \to \mathcal{B}$ demonstrates the superbness of f. To show this, let $A \in \mathcal{A}$ and let $\mathcal{M}$ be a clump in X with $\{A\} \cup \mathcal{M}$ far in X. Since f is initial, $\{fA\} \cup f\mathcal{M}$ is far in Y. Therefore $\{B_A\} \cup f\mathcal{M}$ is far in Y, as required. ∎

[5] "Initial" as used here is equivalent to the category theoretic notion [AHS90; 8.6 and 10.41], and is also equivalent to the formulation in terms of uniform covers in [He88;2.2.1 and 2.2.2]. Finally, it is equivalent to the notion for merotopic spaces that Katětov called *projectively generated* .

[6] "Strict" as used here is equivalent to the term that was defined in terms of uniform covers in [Be92; 4 (5)]. Strict maps were used there (and earlier in [Be77]) to characterize the completion of a nearness space up to uniform isomorphism. See [BH96b] for more details on this.

[7] Exclusive maps were used in [BH79b] to prove that in **Near**, completeness is productive.

[8] A *grill* on X is a nonempty collection of nonempty subsets of X satisfying firstly, that it contains every superset of its members, and secondly, if a finite union of sets is a member, then at least one of those sets is already a member.

Proposition 4.13 *For any nearness space X, the map $e_X : X \to X^*$ sending X into its completion is a superb map.*

Proof. It is not difficult to show that e_X is strict and initial. (A proof can be found in [BH96b; Theorem 10].) ■

Proposition 4.14

(1) Every surjective initial map is exclusive and sleek.

(2) Every exclusive sleek map is superb.

Proof.
(1): Let $f : X \to Y$ be a surjective initial map. Let $\mathcal{G}$ be a near grill on X and let $B \subset Y$ such that $\{f^{-1}B\} \cup \mathcal{G}$ is far in X. Since f is surjective, we have $\{f^{-1}B\} \cup f^{-1}[f\mathcal{G}]$ is far in X and, since f is initial, it follows that $\{B\} \cup f\mathcal{G}$ is far in Y, and therefore f is exclusive. By Proposition 4.12 (1) f is strict, so by Proposition 4.12 (3) f is superb. Therefore, by Proposition 4.6, f is sleek.

(2): Let $f : X \to Y$ be an exclusive map which is also sleek. To show that f is superb, let $\mathcal{B}$ be far in Y, and let $\mathcal{A} = f^{-1}\mathcal{B}$. Then $\mathcal{A}$ is far in X. For each $A \in \mathcal{A}$ choose $B_A \in \mathcal{B}$ such that $A = f^{-1}B_A$. We shall show that the resulting function

$$\mathcal{A} \to \mathcal{B}, \qquad A \mapsto B_A$$

demonstrates the superbness of f. Toward that end, let $\mathcal{M}$ be a clump in X and let $A \in \mathcal{A}$ with $\{A\} \cup \mathcal{M}$ far in X. Let $\mathcal{G} = \text{cluster}(\mathcal{M})$. Then $\{A\} \cup \mathcal{G}$ is far in X, i.e., $\{f^{-1}B_A\} \cup \mathcal{G}$ is far in X. $\mathcal{G}$, being a cluster, is a near grill on X. Since f is exclusive, we get that $\{B_A\} \cup f\mathcal{G}$ is far in Y. Since $\mathcal{M}$ and $\mathcal{G}$ embrace and since f is sleek, it follows that $f\mathcal{M}$ and $f\mathcal{G}$ embrace. Therefore, $\{B_A\} \cup f\mathcal{M}$ is far in Y, as required. ■

5 Completion is Functorial on the Category of Superb Maps

Proposition and Definition 5.1 Composites of superb maps are superb. Uniform isomorphisms are superb. Therefore, there exists the subcategory **Su** of **Near** whose objects are all nearness spaces and whose morphisms are all superb maps.

Proof. Let $f : X \to Y$ and $g : Y \to Z$ be superb maps. Since f and g are excellent, using Proposition 4.7 (3), it is clear that composing a function which demonstrates the superbness of f with a function which demonstrates the superbness of g results in a function which demonstrates the superbness of $g \circ f$. ■

Our primary objective in this section is to show that, when restricted to the subcategory **Su** of **Near**, completion becomes a functor, and is in fact reflective.

Let $\mathbf{ComplSu}_1$ denote the full subcategory of **Su** whose objects are the complete $\mathbf{T}_1$ nearness spaces.

Theorem 5.2 *Let $f : X \to Y$ be a superb map. Then the formula*

$$f^*(\mathcal{G}) = \text{cluster}(f\mathcal{G})$$

defines a superb map $f^* : X^* \to Y^*$ which is the unique continuous[9] *map making the diagram*

$$\begin{array}{ccc} X^* & \xrightarrow{f^*} & Y^* \\ e_X \uparrow & & \uparrow e_Y \\ X & \xrightarrow[f]{} & Y \end{array}$$

commute. In fact, the function $X \mapsto X^*$, $f \mapsto f^*$ is a reflective functor from **Su** *onto* $\mathbf{ComplSu}_1$ *and when restricted to* $\mathbf{Su}_1$*, it is epireflective.*

Proof. If $\mathcal{G}$ is a cluster in X, then it is a clump in X, and since f is excellent, $f\mathcal{G}$ is a clump in Y. Therefore, f^* is a map from $X^* \to Y^*$. We must show it is superb. We first show it is uniformly continuous. To that end, let Λ be far in Y^*. Then there exists $\mathcal{B}$ far in Y such that $\text{cl}_{Y^*} e_Y \mathcal{B}$ corefines Λ. Since f is superb, there exists $\mathcal{A}$ far in X and a function $\mathcal{A} \to \mathcal{B}$, $A \mapsto B_A$ which demonstrates the superbness of f. Then $\text{cl}_{X^*} e_X \mathcal{A}$ is far in X^*, and if we can show that this latter collection corefines $(f^*)^{-1}\Lambda$, it will follow that $(f^*)^{-1}\Lambda$ also is far in X^*, as required. To that end, let $A \in \mathcal{A}$. For some $\lambda \in \Lambda$ we have $\lambda \subset \text{cl}_{Y^*} e_Y B_A$. It suffices to show that $(f^*)^{-1}\lambda \subset \text{cl}_{X^*} e_X A$. So, let $\mathcal{G} \in (f^*)^{-1}\lambda$. Then $f^*\mathcal{G} \in \lambda \subset \text{cl}_{Y^*} e_Y B_A$. Thus $B_A \in f^*\mathcal{G}$ and so $\{B_A\} \cup f\mathcal{G}$ is near. Since $\mathcal{G}$ is a cluster, it is a clump and so $\{A\} \cup \mathcal{G}$ is near. Hence, $A \in \mathcal{G}$ and thus $\mathcal{G} \in \text{cl}_{X^*} e_X A$ and the proof that f^* is uniformly continuous is complete.

If we can show that f^* is excellent, then by Proposition 4.10 it will follow that it is superb. Thus, Let Λ be a clump in X^*; we must show that $f^*\Lambda$ is a clump in Y^*. Let $\Omega = \text{cluster}(\Lambda)$, taken in X^*. Let $\mathcal{M} = \cup\{\cap\lambda \mid \lambda \in \Lambda\}$. By Proposition 3.7 (3), $\mathcal{M}$ is a clump in X. Since f is excellent, $f\mathcal{M}$ is a clump in Y. Let $\Gamma = \text{cluster}(f^*\Lambda)$, taken in Y^* and let $\mathcal{G} = \cup\{\cap\gamma \mid \gamma \in \Gamma\}$. To show that $f^*\Lambda$ is a clump, we must show Γ is near, and for that we need

---

[9] We mean, of course, continuous with respect to the underlying topological structures.

that $\mathcal{G}$ is near, which will follow from

$$\mathcal{G} \subset \text{cluster}(f\mathcal{M}).$$

Toward establishing that inclusion, let $G \in \mathcal{G}$. For some $\gamma \in \Gamma$, we have $G \in \cap\gamma$. So, $\{\gamma\} \cup f^*\Lambda$ is near in Y^*. Hence,

$$(\cap\gamma) \cup (\cup\{\, \cap\beta \mid \beta \in f^*\Lambda \,\})$$

is near in Y. To show that $\{G\} \cup f\mathcal{M}$ is near in Y, as required, it suffices to show that

$$\{G\} \cup f\mathcal{M} \;\subset\; (\cap\gamma) \cup (\cup\{\cap\beta \mid \beta \in f^*\Lambda\,\})$$

and for this we need only show that

$$f\mathcal{M} \subset \cup\{\cap\beta \mid \beta \in f^*\Lambda\,\}.$$

Toward establishing that inclusion, let $M \in \mathcal{M}$. For some $\lambda \in \Lambda$, we have $M \in \cap\lambda$. Let $\beta = f^*\lambda$. $\mathcal{H} \in \lambda$ implies $M \in \mathcal{H}$ which in turn implies $fM \in f\mathcal{H}$. Therefore $fM \in \cap\beta$, as was required, and the proof that f^* is excellent is finished. Therefore, f^* is superb.

To show that the diagram commutes, let $x \in X$. To show that $e_Y(f(x)) = f^*(e_X(x))$, since both sides are clusters, it is sufficient to show that the left side is a subset of the right side. For $A \in e_X(x)$ we have in turn, $x \in \text{cl}A$, $f(x) \in \text{cl}fA$, $fA \in e_Y(f(x))$. Thus, $f[e_X(x)] \subset e_Y(f(x))$. Now for $B \in e_Y(f(x))$ we have $\{B\} \cup f[e_X(x)]$ is near, which means $B \in f^*(e_X(x))$. Hence, $e_Y(f(x)) \subset f^*(e_X(x))$ and the diagram does commute.

Suppose now that in the diagram we replace f^* with a continuous map $g : X^* \to Y^*$ and that the resulting diagram still commutes. We must show that $g = f^*$. Let $\mathcal{G} \in X^*$. For each $A \in \mathcal{G}$ we have $\mathcal{G} \in \text{cl}_{X^*} e_X A$ and, since g is continuous,

$$g(\mathcal{G}) \in \text{cl}_{Y^*} g e_X A = \text{cl}_{Y^*} e_Y f A,$$

which implies that $fA \in g(\mathcal{G})$and then $f\mathcal{G} \subset g(\mathcal{G})$. Therefore, $B \in g(\mathcal{G})$ implies that $\{B\} \cup f\mathcal{G}$ is near and then $B \in f^*(\mathcal{G})$. This shows that $g(\mathcal{G}) \subset f^*(\mathcal{G})$ and since both are clusters, they are equal.

It remains to be shown that completion is a functor on **Su** and is a reflector onto **ComplSu**$_1$.

It follows immediately from Propositions 3.3 and 3.4 (3) that for an identity map $1 : X \to X$ we have that $1^* : X^* \to X^*$ is the identity map. Next, let $f : X \to Y$ and $g : Y \to Z$ be superb maps. We must show that $(g \circ f)^* = g^* \circ f^*$. Toward that end, let $\mathcal{G}$ be a cluster on X. Since $f\mathcal{G} \subset f^*(\mathcal{G}) \in Y^*$, we have

$$gf\mathcal{G} \subset g[f^*(\mathcal{G})] \subset g^*(f^*(\mathcal{G})) \in Z^*.$$

From that, since g and f being excellent implies $gf\mathcal{G}$ is a clump in Z, it follows from Proposition 3.4 (5) that

$$g^*(f^*(\mathcal{G})) = \text{cluster}(gf\mathcal{G}) = (g \circ f)^*(\mathcal{G}).$$

Therefore, we have shown that we have a functor.

In order to show that this functor is reflective, observe first that from Proposition 4.13 we get that $e_X : X \to X^*$ is a superb map. Next, consider the diagram

$$\begin{tikzcd} X \arrow[r, "e_X"] \arrow[dr, "f"'] & X^* \arrow[d, "\overline{f}"] \\ & Y \end{tikzcd}$$

with Y a complete $\mathbf{T}_1$ nearness space and f a superb map. We must come up with a unique superb map $\overline{f}$ making the diagram commute. By [He74c; Proposition 15.9 (1)], since Y is a complete $\mathbf{T}_1$ nearness space, $e_Y : Y \to Y^*$ is a uniform isomorphism. Composing $f^* : X^* \to Y^*$ and $e_Y^{-1} : Y^* \to Y$ we get $\overline{f} = e_Y^{-1} \circ f^*$; since it is a composition of superb maps, it is superb. If $\overline{f}$ were to be replaced by a superb map $u : X^* \to Y$ such that $f = u \circ e_X$, then $e_Y \circ u : X^* \to Y^*$ would be a continuous map such that $(e_Y \circ u) \circ e_X = e_Y \circ f$; hence by the uniqueness result established in the early part of the proof of the present proposition, we have $e_Y \circ u = f^*$ from which it follows that $u = e_Y^{-1} \circ f^* = \overline{f}$.

Finally, we must establish that for a $\mathbf{T}_1$ nearness space X, $e_X : X \to X^*$ is an epimorphism in the category $\mathbf{Su}_1$. For that purpose, let $u, v : X^* \to Y$ be superb maps, Y a $\mathbf{T}_1$ nearness space, such that $u \circ e_X = v \circ e_X$. Then the diagrams

$$\begin{tikzcd} X^* \arrow[r, "e_Y \circ u"] & Y^* \\ X \arrow[u, "e_X"] \arrow[r, "u \circ e_X"'] & Y \arrow[u, "e_Y"'] \end{tikzcd} \qquad \begin{tikzcd} X^* \arrow[r, "e_Y \circ v"] & Y^* \\ X \arrow[u, "e_X"] \arrow[r, "u \circ e_X"'] & Y \arrow[u, "e_Y"'] \end{tikzcd}$$

both commute. Since $e_Y \circ u$ and $e_Y \circ v$ are continuous, by the uniqueness result established earlier in this proof, $e_Y \circ u = e_Y \circ v$. Y being $\mathbf{T}_1$ implies that $e_Y : Y \to Y^*$ is an injective map. Therefore, $u = v$ and the proof is complete. ■

6 An Application to Wallman-type Compactifications

Wallman's construction of a compactification of a $\mathbf{T}_1$ topological space that is referred to by his name was a construction of the space only, without any attention to the question of the extendibility of maps to that compactification. As we mentioned earlier, Arhangel'skii gave a result on extendibility of a map to the Wallman compactification, but Harris gave a more definitive result. The impetus for our study is that result of Harris on how his $\mathcal{WO}$-maps can be used to make the classical Wallman compactification of a $\mathbf{T}_1$ topological space into a functor.

We begin by recalling Harris' definition, and then we proceed through a sequence of generalizations, arriving finally at the superb maps in the setting of separating bases used to generate the so-called Wallman-type compactifications. As we proceed stepwise through the discussion, we will keep relating the ideas to nearness space concepts.

Let $\mathbf{T}_1$ denote the category whose objects are all $\mathbf{T}_1$ topological spaces and whose morphisms are all continuous maps between such. Let $\mathbf{CompT}_1$ denote the full subcategory of $\mathbf{T}_1$ whose objects are the compact $\mathbf{T}_1$ spaces. The classical Wallman compactification gives rise to a function defined on objects $w : \text{obj } \mathbf{T}_1 \to \text{obj } \mathbf{CompT}_1$. Recall that the points of the Wallman compactification were defined by Wallman to be the "ultrafilters" in the lattice of closed sets endowed with the so-called strict topology. From the nearness point of view, the Wallman compactification can be realized as the completion of the contigual reflection of the $\mathbf{T}_1$ topological space X (considered as a topological nearness space). As a completion of a nearness space, X^* has as its points the clusters on X; below in a more general setting we shall indicate the correspondence between clusters and ultrafilters in the lattice of closed sets.

Harris defined the concept of a $\mathcal{WO}$-map $f : X \to Y$ between $\mathbf{T}_1$ topological spaces as follows: If U and V are open subsets of X and Y, respectively, then $U <_f V$ will be written if whenever A is closed in X and $A \subset U$, then $\text{cl}_Y fA \subset V$. If ν is a finite open cover of Y and μ is a finite open cover of X, then $\mu <_f \nu$ will be written if for each $U \in \mu$ there is $V \in \nu$ with $U <_f V$. Then $f : X \to Y$ is defined to be a $\mathcal{WO}$-map iff for each finite open cover ν of Y there is a finite open cover μ of X with $\mu <_f \nu$. The $\mathcal{WO}$-maps are precisely the morphisms of the category $\mathcal{WO}$ of Harris, objects of his category being all $\mathbf{T}_1$ topological spaces. Harris showed that the Wallman compactification $w : \text{obj}\mathbf{T}_1 \to \text{obj}\mathbf{CompT}_1$ thusly gives rise to a functor $w : \mathcal{WO} \to \mathcal{WO} \cap \mathbf{CompT}_1$ and that this functor is an epireflector. Thus, if we denote by $w_X : X \to wX$ the embedding of X into its Wallman compactification, then for any $\mathcal{WO}$-map $f : X \to Y$ from a $\mathbf{T}_1$ space X into a compact $\mathbf{T}_1$ space Y, there exists a unique extension $g : wX \to Y$, i.e., a

$\mathcal{WO}$-map g such that the following diagram is commutative:

$$\begin{tikzcd} X \arrow[r, "w_X"] \arrow[dr, "f"'] & wX \arrow[d, "g"] \\ & Y \end{tikzcd}$$

It is worth noting that, viewing X as a topological nearness space, the relation $\mathrm{cl}_X E \subset \mathrm{interior}_X U$ is merely that $E < U$, i.e., that $\{E, X \setminus U\}$ is far in X (don't confuse this $<$ with the $<_f$ defined above). Thus, in the case that arose above, $A \subset U$, with A closed and U open becomes $A < U$ in X and $\mathrm{cl}_Y fA \subset V$ becomes $fA < V$ in Y. These observations will facilitate understanding the more general situation considered below.

The contribution of Harris can be summarized by the following commutative diagram of object functions and functors (vertical arrows are inclusions):

$$\begin{tikzcd} \mathbf{T}_1 & \mathbf{CompT}_1 \\ \mathcal{WO} \arrow[u] \arrow[r, "\text{Harris}"] & \mathcal{WO} \cap \mathbf{Comp}T_1 \arrow[u] \\ \text{obj } \mathbf{T}_1 \arrow[u] \arrow[r, "\text{Wallman}"] & \text{obj } \mathbf{CompT}_1 \arrow[u] \end{tikzcd}$$

Steiner [St68] applied Wallman's lattice-ultrafilter-space construction to $\mathbf{T}_1$ spaces endowed with a separating base.

Let X be a $\mathbf{T}_1$ topological space and let $\mathcal{L}$ be a base for closed subsets of X. Then $\mathcal{L}$ is called a *separating base* on X provided $\emptyset, X \in \mathcal{L}$, $\mathcal{L}$ is closed under finite unions and finite intersections, and whenever $F \in \mathcal{L}$ and $x \in X \setminus F$, then there exists $E \in \mathcal{L}$ with $x \in E$ and $E \cap (X \setminus F) = \emptyset$. The points of the Wallman-type compactification wX of a $\mathbf{T}_1$ space endowed with a separating base $\mathcal{L}$ are the ultrafilters in the lattice $\mathcal{L}$, given the strict topology.

Generalizing the above mentioned association of a nearness structure (the contigual reflection of X) with the Wallman construction, we also have for a $\mathbf{T}_1$ topological space X endowed with a separating base $\mathcal{L}$ a nearness structure defined as follows: a collection $\mathcal{A}$ of subsets of X is said to be *far* in X provided there exists a finite subset $\mathcal{D}$ of $\mathcal{L}$ that corefines $\mathcal{A}$ and with $\cap\mathcal{D} = \emptyset$. We shall say that *the nearness structure of* X *is determined by the separating base* $\mathcal{L}$. Then the Wallman-type compactification determined by

$\mathcal{L}$ being denoted by wX and the completion of X (considered with this nearness structure) being denoted by X^*, we have the natural one-to-one correspondence

$$\begin{aligned} wX &\longleftrightarrow X^*, \\ \mathcal{U} &\mapsto \mathcal{U} \cap \mathcal{L}, \\ \text{cluster}\mathcal{G} &\leftarrow\!\shortmid \mathcal{G}. \end{aligned}$$

(Observe that every ultrafilter in the lattice $\mathcal{L}$ is a clump in X.)

Let **Sep** denote the category whose objects X are $\mathbf{T}_1$ topological spaces endowed with separating bases $\mathcal{L}_X$ and whose morphisms are functions $f : X \to Y$ which satisfy the following condition: Whenever $\mathcal{G}$ is a finite subset of $\mathcal{L}_Y$ with $\cap\mathcal{G} \neq \emptyset$, then there exists a finite subset $\mathcal{H}$ of $\mathcal{L}_X$ with $\cap\mathcal{H} \neq \emptyset$ such that $\mathcal{H}$ corefines $f^{-1}\mathcal{G}$. Let **CompSep** denote the full subcategory of **Sep** whose objects are all $\mathbf{T}_1$ topological spaces endowed with separating bases with the underlying topology being compact. Steiner's result (although he didn't consider morphisms) can be expressed as: there is an function obj **Sep** $\to$ obj **CompSep** that extends Wallman's construction as in the following commutative diagram of object functions:

$$\begin{array}{ccc} \text{obj } \mathbf{Sep} & \xrightarrow{\text{Steiner}} & \text{obj } \mathbf{CompSep} \\ \uparrow & & \uparrow \\ \text{obj } \mathbf{T}_1 & \xrightarrow{\text{Wallman}} & \text{obj } \mathbf{CompT}_1 \end{array}$$

Analogously to what Harris had done to Wallman's object function, Bentley and Naimpally [BN74a] made Steiner's object function into a functor.

Bentley and Naimpally dualized Harris' definition of $\mathcal{WO}$-map by expressing the property in terms of closed sets and collections of closed sets with empty intersection instead of open sets and open covers.

Thus, the generalization of Harris' $\mathcal{WO}$-maps is as follows. Let $f : X \to Y$ be a function and let X and Y be $\mathbf{T}_1$ topological spaces endowed with separating bases $\mathcal{L}_X$ and $\mathcal{L}_Y$ respectively. If A and B are subsets of X and Y, respectively, then $A <_f B$ will be written if for all $H \subset X$, $\{H, A\}$ is far in X implies $\{fH, B\}$ is far in Y. If $\mathcal{A}$ and $\mathcal{B}$ are finite collections of subsets of X and Y respectively, then $\mathcal{A} <_f \mathcal{B}$ will be written if for each $A \in \mathcal{A}$ there is $B \in \mathcal{B}$ with $A <_f B$. Then $f : X \to Y$ is defined to be a $\mathcal{WO}$-map iff for each finite far collection $\mathcal{B}$ in Y there is a finite far collection $\mathcal{A}$ in X with $\mathcal{A} <_f \mathcal{B}$. There arises the category $\mathcal{WO}$**Sep** whose objects are the same as those of **Sep**, but whose morphisms are all $\mathcal{WO}$-maps between spaces endowed with separating bases.

$\mathcal{WO}$**Sep** is a subcategory, but not a full one, of **Sep**. Furthermore, for $\mathcal{WO}$**Sep**, Steiner's Wallman-type compactification of spaces endowed with a separating base becomes a functor [BN74a] and $\mathcal{WO}$**Sep** $\cap$ **CompSep** is an epireflective subcategory of $\mathcal{WO}$**Sep**. The situation can be summarized by the following commutative diagram of object functions and functors (vertical arrows are inclusions).

$$\begin{array}{ccc}
\mathbf{Sep} & & \mathbf{CompSep} \\
\uparrow & & \uparrow \\
\mathcal{WO}\mathbf{Sep} & \overset{\text{Bentley\&Naimpally}}{\longrightarrow} & \mathcal{WO}\mathbf{Sep} \cap \mathbf{CompSep} \\
\uparrow & & \uparrow \\
\text{obj } \mathbf{Sep} & \overset{\text{Steiner}}{\longrightarrow} & \text{obj } \mathbf{CompSep}
\end{array}$$

The main objective in presenting all the above results on Wallman compactifications and Wallman-type compactifications is to show that Bentley and Naimpally's (and a fortiori Harris') results are corollaries of our extension theorems for superb maps in a certain slightly restricted class of spaces.

We begin with a lemma.

Lemma 6.1 *Let $\mathcal{M}$ be a clump in a nearness space X whose nearness structure is determined by a separating base $\mathcal{L}$ and let $A \subset X$ with $\{A\} \cup \mathcal{M}$ far in X. Then for some $H \subset X$ we have $\{H, A\}$ is far and $\{H\} \cup \mathcal{M}$ is near.*

Proof. Since $\{A\} \cup \mathcal{M}$ is far in X, it is corefined by some finite collection $\mathcal{D} \subset \mathcal{L}$ with $\cap\mathcal{D} = \emptyset$. Let $\mathcal{A} = \{ D \in \mathcal{D} \mid A \not\subset D \}$. Since $\mathcal{A}$ corefines $\mathcal{M}$, it follows that $\mathcal{A} \subset \text{cluster}(\mathcal{M})$. Thus, $\mathcal{A}$ is contained in the $\mathcal{L}$-ultrafilter $\mathcal{L} \cap \text{cluster}(\mathcal{M})$. Therefore, $H = \cap\mathcal{A} \in \text{cluster}(\mathcal{M})$. Clearly, $\mathcal{D}$ corefines $\{H, A\}$ and consequently $\{H, A\}$ is far. ■

In order to proceed, we need a certain strengthening of the result in Lemma 6.1, and to obtain that we restrict the class of spaces with which we are concerned. If a $\mathbf{T}_1$ space X endowed with a separating base $\mathcal{L}$ has the further property that not only are the clumps $\mathcal{M}$ in X micromeric, but the collection

$$\text{stack}_{\mathcal{L}}(\mathcal{M}) = \{ L \in \mathcal{L} \mid M \subset L \text{ for some } M \in \mathcal{M}\}$$

is also micromeric, then we shall say that *clumps in X are tightly covered by $\mathcal{L}$*. For every $\mathbf{T}_1$ space X endowed with a separating base $\mathcal{L}$ which is also a normal base in the sense of Frink [Fr64], clumps in X **are** tightly covered by $\mathcal{L}$ (the easy proof is left to the reader). However, we do not know if every $\mathbf{T}_1$ space endowed with a separating base satisfies this property. For those spaces that do, we have the stronger result:

Lemma 6.2 *Let $\mathcal{M}$ be a clump in a nearness space X whose nearness structure is determined by a separating base $\mathcal{L}$ and for which clumps in X are tightly covered by $\mathcal{L}$. Let $A \subset X$ with $\{A\} \cup \mathcal{M}$ far in X. Then for some $M \in M$ we have $\{M, A\}$ is far.*

Proof. Since $\{A\} \cup \mathcal{M}$ is far in X, it is corefined by some finite collection $\mathcal{D} \subset \mathcal{L}$ with $\cap\mathcal{D} = \emptyset$. Since $\text{stack}_{\mathcal{L}}(\mathcal{M})$ is micromeric, there exist $L \in \text{stack}_{\mathcal{L}}(\mathcal{M})$ and $D \in \mathcal{D}$ with $L \cap D = \emptyset$. There exists $M \in \mathcal{M}$ with $M \subset L$. If $A \not\subset D$, then for some $M' \in \mathcal{M}$, $M' \subset D$ which would imply, first, that $\{M', M\}$ is far and, second, from that, that $\mathcal{M}$ is far. The latter being impossible for a clump, it must be that $A \subset D$. Therefore $\{A, M\}$ is far, as required. ∎

Theorem 6.3 *Let X and Y be $\mathbf{T}_1$ topological spaces endowed with separating bases $\mathcal{L}_X$ and $\mathcal{L}_Y$ respectively (considered also naturally as nearness spaces as defined above) and with X having the property that clumps in X are tightly covered by $\mathcal{L}$. Then every $\mathcal{WO}$-map from X to Y is also a superb map.*

Proof. Let $f : X \to Y$ be a $\mathcal{WO}$-map. To show that f is superb, let $\mathcal{B}$ be far in Y, and without loss of generality, assume that $\mathcal{B}$ is finite. From the fact that f is a $\mathcal{WO}$-map, there exists a finite far collection $\mathcal{A}$ in X with $\mathcal{A} <_f \mathcal{B}$. Using that fact, for each $A \in \mathcal{A}$ choose a $B_A \in \mathcal{B}$ such that $A <_f B_A$. We shall show that the function $\mathcal{A} \to \mathcal{B}$, $A \mapsto B_A$ demonstrates the superbness of f. Toward that end, let $\mathcal{M}$ be a clump in X and let $A \in \mathcal{A}$ with $\{A\} \cup \mathcal{M}$ far. By Lemma 6.2, there exists $M \in \mathcal{M}$ with $\{A, M\}$ far. From that and from the fact that $A <_f B_A$, we get that $\{fM, B_A\}$ is far in Y. Therefore $\{B_A\} \cup f\mathcal{M}$ is far since it is corefined by $\{fM, B_A\}$. ∎

We remark also that the results of Bentley and Naimpally [BN74b] on making Wallman-type $\mathcal{L}$-realcompactifications into a functor is also (in the "tightly covered" case) a consequence of our theorems for superb maps, but we leave the details (which are analogous to those above) to the reader.

7 An Application to Herrlich's Theorem

Recall Herrlich's Theorem (Theorem 1.2 above). We shall show how a slight variant of this theorem is a consequence of our theorems on the extension of superb maps. First we have the following:

Theorem 7.1 *Let X be a nearness space and let A be a dense subspace*[10] *of X such that the inclusion map $A \to Y$ is a strict map. Let Y be a complete $\mathbf{T}_1$ nearness space and let $f : A \to Y$ be a map. Then f has an extension to a superb map $g : X \to Y$ iff $f : A \to Y$ is superb.*

Proof. By Proposition 4.12 (3) the inclusion map, being strict and initial, is superb. If f has a superb extension g, then composing g with the inclusion $A \to X$ gives f as a superb map. For the proof in the converse direction, assume that $f : A \to Y$ is superb. Then by Theorem 5.2, $f^* : A^* \to Y^*$ is superb. Since Y is a complete $\mathbf{T}_1$ nearness space, $e_Y : Y \to Y^*$ is a uniform isomorphism [He74a]. Since A is dense in X and the inclusion is strict, by Theorem 16 in [BH96b] the inclusion $A \to Y$ extends to a uniform isomorphism $u : A^* \to X^*$. Define $g : X \to Y$ by $g = e_Y^{-1} \circ f^* \circ u^{-1} \circ e_X$. Each of the four maps being composed is superb and therefore so is g. Finally, it is obvious that g extends f. ∎

Theorem 7.2 [A Variant of Herrlich's Theorem] *Let X be a nearness space and let A be a dense subspace of X such that the inclusion map $A \to X$ is a strict map. Let Y be a complete $\mathbf{T}_1$ nearness space and let $f : A \to Y$ be a map. Then f can be uniformly continuously extended to X iff whenever $\mathcal{B}$ is far in Y, then $f^{-1}\mathcal{B}$ is far in A.*

Proof. The stated condition is just that the given $f : A \to Y$ be uniformly continuous. Since Y is regular, by Proposition 4.8 this condition is equivalent to the superbness of f. Finally, by Theorem 7.1 the superbness of f is equivalent to the existence of a superb extension, and (Proposition 4.8 again) to a uniformly continuous extension. ∎

References

AHS90 J. Adámek, H. Herrlich, G. E. Strecker. *Abstract and Concrete Categories*, Wiley, New York, 1990.

Be77 H. L. Bentley. Normal nearness spaces, *Quaestiones Math.* **2** (1977), 23–43 and 513.

Be92 H. L. Bentley. Strict extensions of T_1 spaces. In: Recent Developments of General Topology and its Applications, International Conference in Memory of Felix Hausdorff (1868–1942), Berlin (1992), 33–45.

BH78 H. L. Bentley and H. Herrlich. Completion as reflection, *Comment. Math. Univ. Carolinae*, **19** (1978), 541–568.

[10]The statement that A is a subspace of X means that the inclusion map $A \to X$ is initial in **Near**.

BH76 H. L. Bentley and H. Herrlich. Extensions of topological spaces. In: *Topology*, Proc. Memphis Conf., 1975, Marcel Dekker (1976), 129–184.

BH79a H. L. Bentley and H. Herrlich. Completeness for nearness spaces. In: *Topological Structures II, Part 1*, Math. Centre Tracts, **115**, Amsterdam, (1979), 29–40.

BH79b H. L. Bentley and H. Herrlich. Completeness is productive. In: *Categorical Topology*, Lecture Notes in Mathematics **719**, Springer-Verlag, Berlin, (1979), 13–17.

BH96a H. L. Bentley and H. Herrlich. Morita extensions and nearness completions, *Topology and its Appl.*, **82** (1998) 59–65.

BH96b H. L. Bentley and H. Herrlich. On the strict completion of a nearness space. In: *Proceedings of the Eighth Prague Topological Symposium* (1996) 11–21. Download from the internet at http//at.yorku.ca/p/p/a/a/00.htm

BN74a H. L. Bentley and S. A. Naimpally. Wallman $\mathbf{T}_1$ compactifications as epireflections, *General Topology and Appl.*, **4** (1974), 29–41.

BN74b H. L. Bentley and S. A. Naimpally. $\mathcal{L}$-realcompactifications as epireflections, *Proc. Amer. Math. Soc.*, **44** (1974), 196–202.

Da61 A. S. Davis. Indexed systems of neighborhoods for general topological spaces, *Amer. Math. Monthly*, **68** (1961), 886–893.

Fr64 O. Frink. Compactifications and semi-normal spaces, *Amer. J. Math.*, **86** (1964), 602–607.

Ha71 D. Harris. The Wallman compactification as a functor, *General Topology and Appl.*, **1** (1971), 273–281.

He74a H. Herrlich. A concept of nearness, *Gen. Topol. Appl.*, 4 (1974), 191–212.

He74b H. Herrlich. *On the extendibility of continuous functions. General Topology and Appl.*, 4 (1974), 213–215.

He74c H. Herrlich. Topological structures. In: *Topological Structures I, Symposium in Honour of J. de Groot 1914–1972*, Amsterdam 1973, Mathematisch Centrum (1974), 59–122.

He74d H. Herrlich. Topological functors, *General Topology and Appl.*, 4 (1974), 125–142.

He83 H. Herrlich. Categorical topology 1971-1981, In: *General Topology and its Relations to Modern Analysis and Algebra V*, Proceedings of the Fifth Prague Topological Symposium 1981. Edited by J. Novák, Heldermann Verlag, Berlin (1983), 279–383.

He88 H. Herrlich. *Topologie II: Uniforme Räume*, Heldermann Verlag, Berlin, (1988).

II59 V. M. Ivanova and A. A. Ivanov. Contiguity spaces and bicompact extensions of topological spaces, *Dokl. Akad. Nauk SSSR*, **127** (1959), 20–22.

Ka63 M. Katětov. Allgemeine Stetigkeitsstrukturen, In: *Proc. Intern. Congr. Math.*, Stockholm 1962, (1963), 473–479.

Ka65 M. Katětov. On continuity structures and spaces of mappings, *Commentat. Math. Univ. Carol.*, **6** (1965), 257–278.

Mo51 K. Morita. On the simple extension of a space with respect to a uniformity, I–IV, *Proc. Japan Acad.*, **27** (1951), 65–72, 130–137, 166–171, 632–636.

Po64 V. I. Ponomarev. On the extension of multivalued mappings of topological spaces to their compactifications. *Am. Math. Soc. Transl.*, (**2**) **38** (1964), 141–158.

Ša43 N. A. Šanin. On separation in topological spaces, *Dokl. Akad. Nauk SSSR*, **38** (1943), 110–113.

Si00 M. Sioen. On the epireflectivity of the Wallman-Shanin-Type Compactification for approach spaces, *Appl. Categorical Struct.*, **8** (2000), 607–637.

St68 E. F. Steiner. Wallman spaces and compactifications, *Fund. Math.*, **61** (1968), 295–304.

SS73 A. K. Steiner and E. F. Steiner. On semi-uniformities, *Fund. Math.*, **83** (1973), 47–58.

Ta52 A. D. Taĭmanov. On extension of continuous mappings of topological spaces, *Math. Sbornik N. S.*, **31 (73)** (1952), 459–463.

H. L. Bentley
Department of Mathematics
University of Toledo
Toledo, OH 443606 (USA)
`lbentley@math.utoledo.edu`

Characterizations of Subspaces of Important Types of Convergence Spaces in the Realm of Convenient Topology

Gerhard Preuß

ABSTRACT Since in Convenient Topology we are mainly concerned with semiuniform convergence spaces, the question arises how the subspaces of important types of convergence spaces such as topological spaces, pretopological spaces (=closure spaces in the sense of Čech [5]), limit spaces (in the sense of Kowalsky [10] and Fischer [6]) or Kent convergence spaces can be characterized when they are considered as semiuniform convergence spaces (provided all convergence spaces fulfill a certain symmetry condition). This paper presents the solution. Furthermore, the relationships to other important subconstructs of the construct **SUConv** of semiuniform convergence spaces are investigated.

Key words: Semiuniform convergence spaces, filter spaces, sublimit spaces, subpretopological spaces, subtopological spaces, bireflective(resp. bicoreflective) subconstructs.

AMS Subject Classifications (2000): 54A05, 54A20, 54E05, 54E15, 18A40.

0 Introduction

In the realm of Convenient Topology we are mainly concerned with semiuniform convergence spaces and their invariants, i.e., Convenient Topology consists essentially in the study of (full and isomorphism–closed) subconstructs of the construct **SUConv** of semiuniform convergence spaces (cf. [16]). Thus, the study of uniform spaces as well as the study of topological semiuniform convergence spaces (=symmetric topological spaces) belongs to Convenient Topology. Though there is a difference of a "topological" nature between removing a point from the usual topological space $\mathbb{R}_t$ of real numbers and removing a closed interval of length one, the obtained topological spaces are homeomorphic. But if we do the same with the usual uniform space $\mathbb{R}_u$ of real numbers, we obtain non–isomorphic uniform spaces. The reason why uniform spaces behave "well" and topological spaces behave "badly" with respect to the formation of subspaces becomes clear in the framework of semiuniform convergence spaces, namely a subspace of a uniform space is uniform, but a subspace of a topological semiuniform convergence space need not be topological. Thus, the question arises how the 'subspaces of (symmetric) topological spaces' (=subtopological spaces) can be characterized. We solve this question by means of a completion for a certain class of filter spaces (filter spaces have been introduced by Katětov [9]).

The analogous question in the realm of nearness spaces introduced by Herrlich [8] has been solved by Bentley [2]. It turns out that our characterization of subtopological spaces is nothing else than an alternative description of Bentley's subtopological spaces.

By means of a completion for arbitrary filter spaces we solve additionally the problem of characterizing the subspaces of symmetric Kent convergence spaces, symmetric limit spaces and symmetric pretopological spaces and investigate the relationships between all these constructs. Furthermore, it turns out that subtopological spaces have a more comprehensive description than the subspaces of the other above mentioned types of convergence spaces.

The terminology of this article corresponds to [1] and [12].

Conventions.

1) A filter on a set X is not allowed to contain the empty set.
2) Subconstructs are always assumed to be full and isomorphism–closed.

1 Preliminaries

In the following the basic definitions for semiuniform convergence spaces and their invariants are repeated for the convenience of the reader. Those who know them may start with Section 2.

1.1. A *semiuniform convergence space* is a pair $(X, \mathcal{J}_X)$, where X is a set and $\mathcal{J}_X$ a set of filters on $X \times X$ such that the following are satisfied:

UC_1) $\dot{x} \times \dot{x} \in \mathcal{J}_X$ for each $x \in X$, where $\dot{x} = \{A \subset X : x \in A\}$,

UC_2) $\mathcal{G} \in \mathcal{J}_X$ whenever $\mathcal{F} \in \mathcal{J}_X$ and $\mathcal{F} \subset \mathcal{G}$,

UC_3) $\mathcal{F} \in \mathcal{J}_X$ implies $\mathcal{F}^{-1} = \{F^{-1} : F \in \mathcal{F}\} \in \mathcal{J}_X$, where $F^{-1} = \{(y, x) : (x, y) \in F\}$.

A map $f : (X, \mathcal{J}_X) \to (Y, \mathcal{J}_Y)$ between semiuniform convergence spaces is called *uniformly continuous* provided that $(f \times f)(\mathcal{J}_X) \subset \mathcal{J}_Y$. A semiuniform convergence space $(X, \mathcal{J}_X)$ is called a *uniform limit space* provided that the following are satisfied:

UC_4) $\mathcal{F} \in \mathcal{J}_X$ and $\mathcal{G} \in \mathcal{J}_X$ imply $\mathcal{F} \cap \mathcal{G} \in \mathcal{J}_X$,

UC_5) $\mathcal{F} \in \mathcal{J}_X$ and $\mathcal{G} \in \mathcal{J}_X$ imply $\mathcal{F} \circ \mathcal{G} \in \mathcal{J}_X$ (whenever $\mathcal{F} \circ \mathcal{G}$ exists, i.e., $F \circ G = \{(x, y) : \exists z \in X$ with $(x, z) \in G$ and $(z, y) \in F\} \neq \phi$ for every $F \in \mathcal{F}, G \in \mathcal{G}$), where $\mathcal{F} \circ \mathcal{G}$ is the filter generated by the filter base $\{F \circ G : F \in \mathcal{F}$ and $G \in \mathcal{G}\}$.

The construct of all semiuniform convergence spaces (and uniformly continuous maps) is denoted by **SUConv**, whereas **ULim** denotes its subconstruct of all uniform limit spaces. The construct **Unif** of uniform spaces (and uniformly continuous maps) in the sense of A. Weil [17] is bireflectively embedded into **ULim** (cf. e.g., [12;4.2.8.1.]) and **ULim** is a bireflective subconstruct of **SUConv** (cf. [14;2.2]). All these constructs are topological and initial structures are easy to describe, e.g., if X is a set, $(X_i, \mathcal{J}_{X_i})_{i \in I}$ a family of semiuniform convergence spaces and $(f_i : X \to X_i)_{i \in I}$ is a family of maps, then $\mathcal{J}_X = \{\mathcal{F} \in F(X \times X) : (f_i \times f_i)(\mathcal{F}) \in \mathcal{J}_{X_i}$ for each $i \in I\}$ is the initial **SUConv**–structure on X with respect to the given data.

1.2. A *generalized convergence space* is a pair (X, q), where X is a set and $q \subset F(X) \times X$ (provided $F(X)$ denotes the set of all filters on X) such that the following are satisfied:

C_1) $(\dot{x}, x) \in q$ for each $x \in X$,

C_2) $(\mathcal{G}, x) \in q$ whenever $(\mathcal{F}, x) \in q$ and $\mathcal{F} \subset \mathcal{G}$.

If (X, q) is a generalized convergence space, then we say that $\mathcal{F} \in F(X)$ *converges* to $x \in X$ provided that $(\mathcal{F}, x) \in q$. A generalized convergence space (X, q) is called

α) a *Kent convergence space* provided that the following is satisfied:

C_3) $(\mathcal{F} \cap \dot{x}, x) \in q$ whenever $(\mathcal{F}, x) \in q$,

β) a *limit space* provided that the following is satisfied:

C_4) $(\mathcal{F}, x) \in q$ and $(\mathcal{G}, x) \in q$ imply $(\mathcal{F} \cap \mathcal{G}, x) \in q$,

γ) a *pretopological space* provided that the following is satisfied:

C_5) $(\mathcal{U}_q(x), x) \in q$ where $\mathcal{U}_q(x) = \bigcap\{\mathcal{F} \in F(X) : (\mathcal{F}, x) \in q\}$ denotes the *neighborhood filter* of $x \in X$,

δ) a *topological space* provided that it is a pretopological space and for each $x \in X$ the following is satisfied:

C_6) For each $U \in \mathcal{U}_q(x)$ there is some $V \in \mathcal{U}_q(x)$ such that $U \in \mathcal{U}_q(y)$ for all $y \in V$.

ε) *symmetric* provided that the following is satisfied:

S) $(\mathcal{F}, x) \in q$ and $y \in \bigcap_{F \in \mathcal{F}} F$ imply $(\mathcal{F}, y) \in q$.

A map $f : (X, q) \to (X', q')$ between generalized convergence spaces is called *continuous* provided that $(f(\mathcal{F}), f(x)) \in q'$ whenever $(\mathcal{F}, x) \in q$. Every semiuniform convergence space $(X, \mathcal{J}_X)$ has an *underlying (symmetric) Kent convergence space* $(X, q_{\gamma_{\mathcal{J}_X}})$ where $(\mathcal{F}, x) \in q_{\gamma_{\mathcal{J}_X}}$ iff $\mathcal{F} \cap \dot{x} \in \gamma_{\mathcal{J}_X}$ with $\gamma_{\mathcal{J}_X} = \{\mathcal{F} \in F(X) : \mathcal{F} \times \mathcal{F} \in \mathcal{J}_X\}$. A semiuniform convergence $(X, \mathcal{J}_X)$

is called a *convergence space* provided that $\mathcal{J}_X = \{\mathcal{F} \in F(X \times X)$: there are some $x \in X$ and some $\mathcal{G} \in F(X)$ with $(\mathcal{G}, x) \in q_{\gamma_{\mathcal{J}_X}}$ and $\mathcal{F} \supset \mathcal{G} \times \mathcal{G}\}$, i.e., iff it is 'generated' by its convergent filters. The construct **Conv** of convergence spaces (and uniformly continuous maps) is (concretely) isomorphic to the construct $\mathbf{KConv_s}$ of symmetric Kent convergence spaces (and continuous maps) and it is a bicoreflective subconstruct of **SUConv** (the bicoreflector assigns to each semiuniform convergence space its underlying Kent convergence space!). Let $\mathbf{KConv_s}$, $\mathbf{Lim_s}$, $\mathbf{PrTop_s}$ and $\mathbf{Top_s}$ denote the constructs of symmetric Kent convergence spaces, symmetric limit spaces, symmetric pretopological spaces and symmetric topological spaces ($=R_0$–topological spaces) respectively; then each of the constructs in the following list is a bireflective subconstruct of the preceding ones:

$$\mathbf{KConv_s} \supset \mathbf{Lim_s} \supset \mathbf{PrTop_s} \supset \mathbf{Top_s}.$$

For each generalized convergence space (X, q) the *closure* $\overline{A}$ of a subset A of X is defined by $\overline{A} = \{x \in X$: there is some $\mathcal{G} \in F(X)$ with $(\mathcal{G}, x) \in q$ and $A \in \mathcal{G}\}$; furthermore, $O \subset X$ is called *q–open* provided that $X\backslash O = \overline{X\backslash O}$, and $\mathcal{X}_q = \{O \subset X : O$ is q–open$\}$ is a topology on X. Obviously, $O \subset X$ is q–open iff for each $x \in O$ and each $(\mathcal{F}, x) \in q$, $O \in \mathcal{F}$. Additionally, a generalized convergence space (X, q) is topological iff for each $x \in X$, $(\mathcal{U}_{\mathcal{X}_q}(x), x) \in q$, where $\mathcal{U}_{\mathcal{X}_q}(x)$ denotes the neighborhood filter of x in $(X, \mathcal{X}_q)$.

All constructs in this section are topological and initial structures can be easily described, e.g., if X is a set, $(X_i, q_i)_{i \in I}$ a family of symmetric Kent convergence spaces and $(f_i : X \to X_i)_{i \in I}$ a family of maps, then $q = \{(\mathcal{F}, x) \in F(X) \times X : (f_i(\mathcal{F}), f_i(x)) \in q_i$ for each $i \in I\}$ is the initial $\mathbf{KConv_s}$–structure on X with respect to the given data.

1.3. A *filter space* is a pair (X, γ) where X is a set and $\gamma \subset F(X)$ such that the following are satisfied:

F_1) $\dot{x} \in \gamma$ for each $x \in X$,

F_2) $\mathcal{G} \in \gamma$ whenever $\mathcal{F} \in \gamma$ and $\mathcal{F} \subset \mathcal{G}$.

If (X, γ) is a filter space, then the elements of γ are called *Cauchy filters*. A filter space (X, γ) is called a *Cauchy space* provided that the following is satisfied:

F_3) If $\mathcal{F}$ and $\mathcal{G}$ are Cauchy filters such that every member of $\mathcal{F}$ meets every member of $\mathcal{G}$, then $\mathcal{F} \cap \mathcal{G}$ is a Cauchy filter.

A map $f : (X, \gamma) \to (X', \gamma')$ between filter spaces is called *Cauchy continuous* provided that the filter $f(\mathcal{F})$ generated by $\{f[F] : F \in \mathcal{F}\}$ belongs to γ' for each $\mathcal{F} \in \gamma$.

For every semiuniform convergence space $(X, \mathcal{J}_X)$, $(X, \gamma_{\mathcal{J}_X})$ is called the *underlying filter space.*

The construct **Fil** of filter spaces (and Cauchy continuous maps) is topological and initial structures can be easily described; namely if X is a set, $(X_i, \gamma_i)_{i \in I}$ a family of filter spaces and $(f_i : X \to X_i)_{i \in I}$ a family of maps, then $\gamma = \{\mathcal{F} \in F(X) : f_i(\mathcal{F}) \in \gamma_i$ for each $i \in I\}$ is the initial **Fil**–structure on X with respect to the given data. The subconstruct **Chy** of all Cauchy spaces is bireflective in **Fil** (cf. [13;3.1.]).

Every filter space (X, γ) has an *underlying symmetric Kent convergence space* (X, q_γ), where $(\mathcal{F}, x) \in q_\gamma$ iff $\mathcal{F} \cap \dot{x} \in \gamma$. A filter space (X, γ) is called *complete* provided that each $\mathcal{F} \in \gamma$ converges in (X, q_γ). Obviously, the construct **CFil** of complete filter spaces (and Cauchy continuous maps) is (concretely) isomorphic to $\mathbf{KConv_s}$ and bicoreflective in **Fil** (the bicoreflector assigns to each filter space its underlying symmetric Kent convergence space!). A semiuniform convergence space $(X, \mathcal{J}_X)$ is called **Fil**–*determined* provided that $\mathcal{J}_X = \{\mathcal{F} \in F(X \times X)$: there is some $\mathcal{G} \in \gamma_{\mathcal{J}_X}$ with $\mathcal{G} \times \mathcal{G} \subset \mathcal{F}\}$, i.e., iff it is 'generated' by its Cauchy filters. The construct **Fil-D-SUConv** of all **Fil**–determined semiuniform convergence spaces (and uniformly continuous maps) is (concretely) isomorphic to **Fil** and it is bireflective and bicoreflective in **SUConv** (cf. [14]); the bicoreflector assigns to each semiuniform convergence space its underlying filter space.

1.4. A semiuniform convergence space $(X, \mathcal{J}_X)$ is called *topological* provided it is a convergence space and the corresponding symmetric Kent convergence space is topological. It is easily checked that the construct **T-SUConv** of topological semiuniform convergence spaces (and uniformly continuous maps) is (concretely) isomorphic to $\mathbf{Top_s}$ and that it is closed under formation of products and closed (!) subspaces in **SUConv**. Arbitrary subspaces, formed in **SUConv**, of topological semiuniform convergence spaces need not be topological, e.g., if $X = \mathbb{R} \setminus \{0\}$, then the (open) subspace $(X, \mathcal{J}_X)$ (in **SUConv**) of the usual topological space $\mathbb{R}_t$ of real numbers (regarded as a semiuniform convergence space) is not topological since it is not complete.

2 A Filter Space Completion and Its Applications

2.1 Theorem. *Let (X, γ) be a filter space. Put $\hat{X} = X \cup \{\mathcal{F} \in \gamma$: $\mathcal{F}$ is non–convergent in $(X, q_\gamma)\}$ and $\hat{\gamma} = \{\mathcal{H} \in F(\hat{X}) : \mathcal{H} \supset \mathcal{G}$ for some $\mathcal{G} \in \hat{\mathcal{B}}\}$ where $\hat{\mathcal{B}} = \{i(\mathcal{F}) \cap \dot{x} : (\mathcal{F}, x) \in q_\gamma\} \cup \{i(\mathcal{F}) \cap \dot{\mathcal{F}} : \mathcal{F} \in \gamma$ is non–convergent in $(X, q_\gamma)\}$*) and $i : X \to \hat{X}$ denotes the inclusion map. Then $(\hat{X}, \hat{\gamma})$ is a complete filter space containing (X, γ) as a dense subspace.*

---

*) If A is a subset of a set Y and $x \in A$, then we do not make a notational distinction between the ultrafilter $\dot{x}$ formed in A and the ultrafilter $\dot{x}$ formed in Y; usually, $\mathcal{F} \cap \dot{x}$ means that $\dot{x}$ is formed in A provided that $\mathcal{F}$ is a filter on A and $\mathcal{F} \cap \dot{x}$ means that $\dot{x}$ is formed in Y provided that $\mathcal{F}$ is filter on Y.

Proof.

0) Obviously, $(\hat{X},\hat{\gamma})$ is a filter space.

1) (X,γ) is a subspace of $(\hat{X},\hat{\gamma})$, i.e., $\gamma = \{\mathcal{F} \in F(X) : i(\mathcal{F}) \in \hat{\gamma}\}$: If $\mathcal{F} \in \gamma$, then $i(\mathcal{F}) \in \hat{\gamma}$. Conversely, let $\mathcal{F} \in F(X)$ such that $i(\mathcal{F}) \in \hat{\gamma}$. Then there is some $(\mathcal{G},x) \in q_\gamma$ with $i(\mathcal{F}) \supset i(\mathcal{G}) \cap \dot{x}$ or there is some non–convergent $\mathcal{G} \in \gamma$ with $i(\mathcal{F}) \supset i(\mathcal{G}) \cap \dot{\mathcal{G}}$. In the first case $\mathcal{G}' = \mathcal{G} \cap \dot{x} \in \gamma$ and $i(\mathcal{G}') = i(\mathcal{G}) \cap i(\dot{x}) = i(\mathcal{G}) \cap \dot{x} \subset i(\mathcal{F})$; thus, $\mathcal{F} = i^{-1}(i(\mathcal{F})) \supset i^{-1}(i(\mathcal{G}')) = \mathcal{G}'$ and consequently, $\mathcal{F} \in \gamma$. In the second case $\mathcal{F} = i^{-1}(i(\mathcal{F})) \supset i^{-1}(i(\mathcal{G}) \cap \dot{\mathcal{G}}) = \mathcal{G}$ which implies $\mathcal{F} \in \gamma$.

2) X is dense in $(\hat{X},\hat{\gamma})$, i.e., $\hat{X} \subset \overline{X} = \{y \in \hat{X}$: there is some $\mathcal{G} \in F(\hat{X})$ with $(\mathcal{G},y) \in q_{\hat{\gamma}}$ and $X \in \mathcal{G}\}$: If $y \in \hat{X}$, then $y \in X$ or $y = \mathcal{F} \in \gamma$ is non–convergent. In the first case $y \in \overline{X}$, since $X \in i(\dot{y}) \xrightarrow{q_{\hat{\gamma}}} i(y) = y$. In the second case $y \in \overline{X}$, since $X \in i(\mathcal{F}) \xrightarrow{q_{\hat{\gamma}}} \mathcal{F} = y$.

3) $(\hat{X},\hat{\gamma})$ is complete: If $\mathcal{G} \in \hat{\gamma}$, then $\mathcal{G} \supset i(\mathcal{F}) \cap \dot{x}$ with $(\mathcal{F},x) \in q_\gamma$ or $\mathcal{G} \supset i(\mathcal{F}) \cap \dot{\mathcal{F}}$ where $\mathcal{F} \in \gamma$ is non–convergent. In the first case it follows from $i(\mathcal{F}) \cap \dot{x} \xrightarrow{q_{\hat{\gamma}}} x$ that $\mathcal{G} \xrightarrow{q_{\hat{\gamma}}} x$. In the second case it follows that $\mathcal{G} \cap \dot{\mathcal{F}} \in \hat{\gamma}$, i.e., $\mathcal{G} \xrightarrow{q_{\hat{\gamma}}} \mathcal{F}$. ■

2.2 Proposition. *Let (X,γ) be a filter space and $(\hat{X},\hat{\gamma})$ its above completion. If $(\hat{X},q_{\hat{\gamma}})$ is the underlying Kent convergence space of $(\hat{X},\hat{\gamma})$, then the following are valid:*

1. *$x \in X$: $(\hat{\mathcal{F}},x) \in q_{\hat{\gamma}} \iff \hat{\mathcal{F}} \cap \dot{x} \supset \mathcal{G}$ for some $\mathcal{G} \in \gamma$,*
2. *$\mathcal{F} \in \hat{X} \backslash X$: $(\hat{\mathcal{F}},\mathcal{F}) \in q_{\hat{\gamma}} \iff \hat{\mathcal{F}} \supset i(\mathcal{F}) \cap \dot{\mathcal{F}}$.*

Proof.

1. " $\Rightarrow$". The case $\hat{\mathcal{F}} \cap \dot{x} \supset i(\mathcal{F}) \cap \dot{\mathcal{F}}$ for some non–convergent $\mathcal{F} \in \gamma$ cannot occur since otherwise for each $F \in \mathcal{F}$ there were some $\hat{F} \in \hat{\mathcal{F}}$ such that $\hat{F} \cup \{x\} \subset F \cup \{\mathcal{F}\}$, and consequently, $x \in F$, i.e., $\mathcal{F} \subset \dot{x}$ which would imply $\mathcal{F} \cap \dot{x} = \mathcal{F} \in \gamma$, i.e., $\mathcal{F} \xrightarrow{q_\gamma} x$ – a contradiction. If $\hat{\mathcal{F}} \cap \dot{x} \supset i(\mathcal{F}) \cap \dot{y}$ for some $(\mathcal{F},y) \in q$, then $\hat{\mathcal{F}} \cap \dot{x} \supset \mathcal{G}$ for $\mathcal{G} = \mathcal{F} \cap \dot{y} \in \gamma$.

2. " $\Rightarrow$". The case $\hat{\mathcal{F}} \cap \dot{\mathcal{F}} \supset i(\mathcal{G}) \cap \dot{x}$ for some $(\mathcal{G},x) \in q_\gamma$ cannot occur, since $X \in \mathcal{G}$ and thus $X \cup \{x\} = X \in i(\mathcal{G}) \cap \dot{x}$ but on the other side $X \notin \hat{\mathcal{F}} \cap \dot{\mathcal{F}}$, i.e., $X \not\supset \hat{F} \cup \{\mathcal{F}\}$ for each $\hat{F} \in \hat{\mathcal{F}}$, because $\mathcal{F} \notin X$. Consequently, it remains the case that there is some non–convergent $\mathcal{G} \in \gamma$ such that $\hat{\mathcal{F}} \cap \dot{\mathcal{F}} \supset i(\mathcal{G}) \cap \dot{\mathcal{G}}$. Since $X \cup \{\mathcal{G}\} \in i(\mathcal{G}) \cap \dot{\mathcal{G}}$, one obtains $X \cup \{\mathcal{G}\} = \hat{F} \cup \{\mathcal{F}\}$ for some $\hat{F} \in \hat{\mathcal{F}}$. Thus, $\mathcal{F} = \mathcal{G}$, since $\mathcal{F} \notin X$. Hence, $\hat{\mathcal{F}} \cap \dot{\mathcal{F}} \supset i(\mathcal{F}) \cap \dot{\mathcal{F}}$, i.e., $\hat{\mathcal{F}} \supset i(\mathcal{F}) \cap \dot{\mathcal{F}}$. ■

The inverse implications in both cases are trivial.

2.3 Definitions.

1) A filter space (X,γ) is called

a) a *sublimit space* provided that (X,q_γ) is a limit space,

b) *subpretopological* provided that (X,q_γ) is a pretopological space,

c) *weakly subtopological* provided that (X,q_γ) is a topological space.

2) **SubLim**, **SubPrTop** and $\mathbf{SubTop_w}$ denote the constructs of sublimit spaces, subpretopological and weakly subtopological filter spaces (and Cauchy continuous maps) respectively.

2.4 Proposition. *Each of the constructs in the following list is a bireflective subconstruct of the preceding ones:*

$$\mathbf{Fil} \supset \mathbf{SubLim} \supset \mathbf{SubPrTop} \supset \mathbf{SubTop_w}$$

Proof. Let $\mathcal{A}$ be the construct **SubLim**, **SubPrTop** or $\mathbf{SubTop_w}$ respectively and $\mathcal{B}$ the construct $\mathbf{Lim_s}$, $\mathbf{PrTop_s}$ or $\mathbf{Top_s}$ respectively. It suffices to prove that $\mathcal{A}$ is bireflective in **Fil**. Let $(f_i : (X, \gamma) \to (X_i, \gamma_i))_{i \in I}$ be an initial source in **Fil** such that $(X_i, \gamma_i) \in |\mathcal{A}|$ for each $i \in I$. Then $(f_i : (X, q_\gamma) \to (X, q_{\gamma_i}))_{i \in I}$ is an initial source in $\mathbf{KConv_s}$. Since $\mathcal{B}$ is bireflective in $\mathbf{KConv_s}$ and by assumption all (X_i, q_{γ_i}) belong to $|\mathcal{B}|$ it follows that $(X, q_\gamma) \in |\mathcal{B}|$, i.e., $(X, \gamma) \in |\mathcal{A}|$. ■

2.5 Definitions.

1) A filter space (X, γ) is called
   a) a *limit filter space* provided that it is a complete sublimit space,
   b) *pretopological* provided that it is complete and subpretopological,
   c) *topological* provided that it is complete and weakly subtopological.

2) **Lim-Fil**, **PrTop-Fil** and **Top-Fil** denote the constructs of limit filter spaces, pretopological and topological filter spaces (and Cauchy continuous maps) respectively.

2.6 Remark. The following isomorphisms are concrete:

1. **Lim-Fil** $\cong$ $\mathbf{Lim_s}$
2. **PrTop-Fil** $\cong$ $\mathbf{PrTop_s}$
3. **Top-Fil** $\cong$ $\mathbf{Top_s}$

(Note: **CFil** $\cong$ $\mathbf{KConv_s}$ [cf. Section 1.3]).

2.7 Theorem. *Let (X, γ) be a filter space.*

1) *The following are equivalent:*
   (a) (X, γ) is a sublimit space.
   (b) $(\hat{X}, \hat{\gamma})$ is a sublimit space.
   (c) $(\hat{X}, \hat{\gamma})$ is a limit filter space.
   (d) (X, γ) is a dense subspace (in **Fil**) *of some limit filter space.*
   (e) (X, γ) is a subspace (in **Fil**) *of some limit filter space.*
   (f) $(X, \gamma) (= (X, \mathcal{J}_\gamma))$ is a dense subspace (in **SUConv**) *of some (symmetric) limit space (Y, q) $(= (Y, \mathcal{J}_{\gamma_q}))$.*
   (g) $(X, \gamma) (= (X, \mathcal{J}_\gamma))$ is a subspace (in **SUConv**) *of some (symmetric) limit space $(Y, q) (= (Y, \mathcal{J}_{\gamma_q}))$.*

2) *The following are equivalent:*
   (a) (X, γ) is subpretopological.
   (b) $(\hat{X}, \hat{\gamma})$ is subpretopological.
   (c) $(\hat{X}, \hat{\gamma})$ is pretopological.

(d) (X, γ) *is a dense subspace (in* **Fil***) of some pretopological filter space.*

(e) (X, γ) *is a subspace (in* **Fil***) of some pretopological filter space.*

(f) $(X, \gamma)(= (X, \mathcal{J}_\gamma))$ *is a dense subspace (in* **SUConv***) of some (symmetric) pretopological space* $(Y, q)(= (Y, \mathcal{J}_{\gamma_q}))$,

(g) $(X, \gamma)(= (X, \mathcal{J}_\gamma))$ *is a subspace (in* **SUConv***) of some (symmetric) pretopological space* $(Y, q)(= (Y, \mathcal{J}_{\gamma_q}))$.

Proof. Since 1) and 2) can be proved similarly, we restrict ourselves to the proof of 2):

(a) $\Rightarrow$ (b). Let $y \in \hat{X}$. Then $y \in X$ or $y \in \hat{X} \setminus X$. In either case we have to prove that $(\mathcal{U}_{q_{\hat{\gamma}}}(y), y) \in q_{\hat{\gamma}}$, where $\mathcal{U}_{q_{\hat{\gamma}}}(y) = \bigcap\{\hat{\mathcal{F}} \in F(\hat{X}) : (\hat{\mathcal{F}}, y) \in q_{\hat{\gamma}}\}$. This follows in the first case from 2.2.1. and in the second case from 2.2.2.

(b) $\Rightarrow$ (c). Since $(\hat{X}, \hat{\gamma})$ is complete, this implication is obvious.

(c) $\Rightarrow$ (d). Obvious.

(d) $\Rightarrow$ (e). Obvious.

(e) $\Rightarrow$ (a). This implication follows from the fact that every pretopological filter space is subpretopological and that by Proposition 2.4. **SubPrTop** is closed under formation of subspaces in **Fil**.

Furthermore, (d) $\Leftrightarrow$ (f) (resp. (e) $\Leftrightarrow$ (g)), since subspaces in **Fil** are formed as in **SUConv** (note: **Fil** is bireflectively embedded in **SU-Conv**). ■

2.8 Remarks.

1) It follows also from Theorem 2.1. that *the construct of all subspaces (in* **Fil** *[resp.* **SUConv***]) of symmetric Kent convergence spaces is* **Fil** *itself* (cf. also [15;3.6]).

2) **SubLim** has also been studied by Bentley, Herrlich and Lowen–Colebunders [3] in the realm of merotopic spaces; it has been denoted there by **C**. In particular, they proved that it is not cartesian closed. Furthermore, it contains **Chy** as a bireflective subconstruct.

3 Subtopological Spaces

3.1 Definition. A filter space (X, γ) is called *subtopological* provided that each $\mathcal{F} \in \gamma$ contains some $\mathcal{G} \in \gamma$ with a q_γ–open base $\mathcal{B}$ (i.e., for each $B \in \mathcal{B}, B \in \mathcal{X}_{q_\gamma}$).

3.2 Proposition. *Let* (X, γ) *be a filter space. Then the following are equivalent:*

(1) (X, γ) *is subtopological,*

(2) *Each* $\mathcal{F} \in \gamma$ *contains some* $\mathcal{G} \in \gamma$ *such that for each* $G \in \mathcal{G}, int_{\mathcal{X}_{q_\gamma}} G \in \mathcal{G}$,

(3) *For each Cauchy filter* $\mathcal{F}$ *in* (X, γ) *the topological neighborhood filter*

of $\mathcal{F}$, denoted by $\mathcal{U}_{\mathcal{X}_{q_\gamma}}(\mathcal{F})$, is a Cauchy filter in (X,γ), where $\mathcal{U}_{\mathcal{X}_{q_\gamma}}(\mathcal{F}) = \{U \subset X : U \supset O_F \supset F$ for some $F \in \mathcal{F}$ and some $O_F \in \mathcal{X}_{q_\gamma}\}$.

3.3 Proposition. *Every subtopological filter space (X,γ) is weakly subtopological.*

Proof. Obvious, since $\mathcal{U}_{\mathcal{X}_{q_\gamma}}(\dot{x}) = \mathcal{U}_{\mathcal{X}_{q_\gamma}}(x) = \mathcal{U}_{\mathcal{X}_{q_\gamma}} \cap \dot{x} \in \gamma$. ■

3.4 Remark. *A weakly subtopological filter space need not be subtopological* as the following example shows: Let X be the set $\mathbb{R}\setminus\{0\}$ and $\mathcal{X}$ the topology induced by the usual topology on $\mathbb{R}$. Put $\gamma = \gamma_{q_\mathcal{X}} \cup \{\mathcal{F} \in F(X) : \mathcal{F} \supset \mathcal{G}\}$ where $\mathcal{G}$ denotes the elementary filter of the sequence $(1/n)$ in X. Then (X,γ) is a filter space whose underlying (symmetric) Kent convergence space is the topological space $(X, q_\mathcal{X})$. Thus, (X,γ) is weakly subtopological, but not subtopological, since $\mathcal{G} \in \gamma$ does not have a $q_\mathcal{X}$–open base.

3.5 Proposition. *The construct* **SubTop** *of all subtopological filter spaces (and Cauchy continuous maps) is a bireflective subconstruct of* **Fil**.

Proof.

1) **SubTop** is closed under formation of subspaces in **Fil**: Let $(X',\gamma') \in |\mathbf{SubTop}|$ and (X,γ) a subspace in **Fil**, i.e., $X \subset X'$ and $\gamma = \{\mathcal{F} \in F(X) : i(\mathcal{F}) \in \gamma'\}$, where $i : X \to X'$ denotes the inclusion map. By assumption, for each $\mathcal{F} \in \gamma$, there is some $\mathcal{G} \in \gamma'$ such that $\mathcal{G} \subset i(\mathcal{F})$ and $\mathcal{G}$ has a $q_{\gamma'}$–open base. Thus, $\mathcal{F} \supset i^{-1}(\mathcal{G})$, where $i^{-1}(\mathcal{G})$ is a Cauchy filter in (X,γ) with a q_γ–open base ($i : (X,q_\gamma) \to (X',q'_\gamma)$ as well as $i : (X, q_{\mathcal{X}_{q_\gamma}}) \to (X', q_{\mathcal{X}_{q_{\gamma'}}})$ are continuous, i.e., $i : (X, \mathcal{X}_{q_\gamma}) \to (X', \mathcal{X}_{q_{\gamma'}})$ is continuous!)

2) **SubTop** is closed under formation of products in **Fil**. Let $((X_i,\gamma_i))_{i\in I}$ be a family of **SubTop**–objects and $((\prod_{i\in I} X_i, \gamma), (p_i)_{i\in I})$ their product in **Fil**, i.e., $\gamma = \{\mathcal{F} \in F(\Pi X_i) : p_i(\mathcal{F}) \in \gamma_i$ for each $i \in I\}$, where $p_i : \Pi X_i \to X_i$ denotes the i–th projection. If $\mathcal{F} \in \gamma$, then by assumption, for each $i \in I$, there is some $\mathcal{G}_i \in \gamma_i$ such that $\mathcal{G}_i \subset p_i(\mathcal{F})$ and $\mathcal{G}_i$ contains $int_{\mathcal{X}_{q_{\gamma_i}}} G_i$ for each $G_i \in \mathcal{F}_i$. Hence, the product filter $\prod_{i\in I} \mathcal{F}_i \subset \mathcal{F}$ contains the product filter $\prod_{i\in I} \mathcal{G}_i \in \gamma$ (note: $p_i[\Pi \mathcal{G}_i] = \mathcal{G}_i \in \gamma_i$ for each $i \in I$) which has a q_γ–open base $\mathcal{B} = \{\prod_{i\in I} int_{\mathcal{X}_{q_{\gamma_i}}} G_i : G_i \in \mathcal{G}_i$ for each $i \in I$ and $G_i \neq X_i$ for at most finitely many $i \in I\}$ (note: $\prod_{i\in I} int_{\mathcal{X}_{q_{\gamma_i}}} G_i \in \mathcal{X}$ provided that $(\prod_{i\in I} X_i, \mathcal{X})$ denotes the product of $((X_i, \mathcal{X}_{q_{\gamma_i}}))_{i\in I}$ in $\mathbf{Top_s}$.)

3) All indiscrete **Fil**–objects, i.e., all filter spaces (X,γ) such that $\gamma = F(X)$, belong to $|\mathbf{SubTop}|$, since for each $\mathcal{F} \in F(X), \mathcal{U}_{\mathcal{X}_{q_\gamma}}(\mathcal{F}) \in F(X)$. ■

It follows from 1), 2) and 3) that **SubTop** is bireflective in **Fil** (cf. [12;2.2.11.②]).

3.6 Theorem. *Let (X,γ) be a subtopological filter space. Put $X^* = X \cup \{\mathcal{F} \in \gamma : \mathcal{F}$ has a q_γ–open base and is non–convergent$\}$ and $\gamma^* = \{\mathcal{H} \in F(X^*) : \mathcal{H} \supset \mathcal{G}$ for some $\mathcal{G} \in \mathcal{B}^*\}$ where $\mathcal{B}^* = \{i(\mathcal{F}) \cap \dot{x} : (\mathcal{F},x) \in q_\gamma\} \cup \{i(\mathcal{F}) \cap \dot{\mathcal{F}} : \mathcal{F} \in \gamma$ has a q_γ–open base and is non–convergent$\}$ and $i : X \to X^*$ denotes the inclusion map. Then (X^*,γ^*) is a complete filter space containing (X,γ) as a dense subspace.*

The *proof* of Proposition 3.5. is omitted, since it is similar to Theorem 2.1.

3.7 Proposition. *Let (X,γ) be a subtopological filter space and (X^*,γ^*) its above completion (cf. 3.6.). If (X^*,q_{γ^*}) is the underlying Kent convergence space of (X^*,γ^*), then the following is valid:*

1. *$x \in X : (\mathcal{F}^*,x) \in q_{\gamma^*} \iff \mathcal{F}^* \cap \dot{x} \supset \mathcal{G}$ for some $\mathcal{G} \in \gamma$.*
2. *$\mathcal{F} \in X^* \backslash X : (\mathcal{F}^*,\mathcal{F}) \in q_{\gamma^*} \iff \mathcal{F}^* \supset i(\mathcal{F}) \cap \dot{\mathcal{F}}$.*

Proof. Proposition 3.7. is proved analogously to Proposition 2.2. ■

3.8 Corollary. *Let (X,γ) be a subtopological filter space and (X^*,γ^*) its above completion. The neighborhood filters of the points of (X^*,q_{γ^*}) are obtained as follows:*

1. *$x \in X : \mathcal{U}_{q_{\gamma^*}}(x) = i(\mathcal{U}_{q_\gamma}(x))$,*
2. *$\mathcal{F} \in X^* \backslash X : \mathcal{U}_{q_{\gamma^*}}(\mathcal{F}) = i(\mathcal{F}) \cap \dot{\mathcal{F}}$.*

Proof. Apply Proposition 3.7. ■

3.9 Proposition. *Every topological filter space is subtopological.*

Proof. Let (X,γ) be a topological filter space. Then (X,γ) is weakly subtopological, i.e., (X,q_γ) is topological, and complete, i.e., from $\mathcal{F} \in \gamma$ follows the existence of some $x \in X$ with $\mathcal{F} \xrightarrow{q_\gamma} x$ or equivalently $\mathcal{F} \supset \mathcal{U}_{q_\gamma}(x)$, where $\mathcal{U}_{q_\gamma}(x)$ has a q_γ–open base. Thus, (X,γ) is also subtopological. ■

3.10 Theorem. *Let (X,γ) be a filter space. Then the following are equivalent:*

(1) (X,γ) is subtopological.
(2) (X,γ) is a dense subspace(in **Fil***) of some topological filter space.*
(3) (X,γ) is a subspace(in **Fil***) of some topological filter space.*

Proof.
(1) $\Rightarrow$ (2). It suffices to show that the completion (X^*,γ^*) of (X,γ) (cf. Theorem 3.6.) is a topological filter space, i.e., (X^*,q_{γ^*}) is topological. Since by Corollary 3.8., C_5) is valid, C_6) must be proved for (X^*,q_{γ^*}).

1. $x \in X$: If $U_x^* \in \mathcal{U}_{q_{\gamma^*}}(x)$, then, by Corollary 3.8.1., there is some $U_x \in \mathcal{U}_{q_\gamma}(x)$ with $U_x \subset U_x^*$. Since (X,q_γ) is topological, there is some $V_x \in \mathcal{U}_{q_\gamma}(x) \subset \mathcal{U}_{q_{\gamma^*}}(x)$ such that for each $z \in V_x$, $U_x \in \mathcal{U}_{q_\gamma}(z) \subset i(\mathcal{U}_{q_\gamma}(z)) = \mathcal{U}_{q_{\gamma^*}}(z)$.

2. $\mathcal{F} \in X^*\backslash X$: Let $U^*_{\mathcal{F}} \in \mathcal{U}_{q_{\gamma^*}}(\mathcal{F}) = i(\mathcal{F}) \cap \dot{\mathcal{F}}$ (cf. Corollary 3.8.2.), i.e., $U^*_{\mathcal{F}} \supset F \cup \{\mathcal{F}\}$ for some $F \in \mathcal{F}$. Put $V^*_{\mathcal{F}} = (int_{\mathcal{X}_{q_\gamma}} F) \cup \{\mathcal{F}\}$. Then $V^*_{\mathcal{F}} \subset U^*_{\mathcal{F}}$ and since by assumption, $int_{\mathcal{X}_{q_\gamma}} F \in \mathcal{F}, V^*_{\mathcal{F}} \in \mathcal{U}_{q_{\gamma^*}}(\mathcal{F})$. Let $z \in V^*_{\mathcal{F}}$. Then a) $z = \mathcal{F}$ or b) $z \in int_{\mathcal{X}_{q_\gamma}} F$. In case a), $U^*_{\mathcal{F}} \in \mathcal{U}_{q_{\gamma^*}}(\mathcal{F})$ by assumption. In case b), $U^*_{\mathcal{F}} \in \mathcal{U}_{q_{\gamma^*}}(z)$ since $int_{\mathcal{X}_{q_\gamma}} F \in \mathcal{U}_{q_\gamma}(z) \subset i(\mathcal{U}_{q_\gamma}(z)) = \mathcal{U}_{q_{\gamma^*}}(z)$ and $U^*_{\mathcal{F}} \supset int_{\mathcal{X}_{q_\gamma}} F$.

(2) $\Rightarrow$ (3). Obvious.

(3) $\Rightarrow$ (1). Let (Y, η) be a topological filter space containing (X, γ) as a subspace. Since every topological filter space is subtopological (cf. Proposition 3.9.) and **SubTop** is bireflective in **Fil** (and thus closed under formation of subspaces), (X, γ) is also subtopological. ■

3.11 Definition. A semiuniform convergence space $(X, \mathcal{J}_X)$ is called *subtopological* provided that it is **Fil**–determined and its corresponding filter space $(X, \gamma_{\mathcal{J}_X})$ is subtopological.

3.12 Corollary. *Let* $(X, \mathcal{J}_X)$ *be a semiuniform convergence space. Then the following are equivalent:*

(1) $(X, \mathcal{J}_X)$ *is subtopological.*

(2) $(X, \mathcal{J}_X)$ *is a subspace (in* **SUConv**) *of some topological semiuniform convergence space.*

(3) $(X, \mathcal{J}_X)$ *is a dense subspace (in* **SUConv**) *of some topological semiuniform convergence space.*

Proof. Since **Fil-D-SUConv** $\cong$ **Fil** is bireflective in **SUConv** (and thus subspaces in **Fil** are formed as in **SUConv**) and the construct **T-SUConv** of all topological semiuniform convergence spaces (and uniformly continuous maps) is (concretely) isomorphic to $\mathbf{Top_s}$ as well as to **Top-Fil**, the above corollary is an immediate consequence of 3.10. ■

3.13 Remarks.

1) Obviously, the construct of all subspaces (in **SUConv**) of topological semiuniform convergence spaces (and uniformly continuous maps) is (concretely) isomorphic to **SubTop**.

2) Bentley [2] defines 'subtopological spaces' to be subspaces (in the construct **Near** of nearness spaces) of topological nearness spaces (= symmetric topological spaces) and shows that they are those filter spaces (= filtermerotopic spaces) which are nearness spaces. Since there is no difference in forming subspaces of symmetric topological spaces in **Near**, **Mer** (= construct of merotopic spaces in the sense of Katětov [9]) or **Fil** respectively, it follows from Theorem 3.10. that a *filter space (= filtermerotopic space) is subtopological iff it is a nearness space.*

3.14 Definition. A Cauchy space (X, γ) is called *uniformizable* provided that there is a uniformity $\mathcal{W}$ on X such that γ is the set of all Cauchy filters in $(X, \mathcal{W})$ (the uniformity $\mathcal{W}$ is called a *compatible uniformity*).

3.15 Remark. The construct **UChy** of all uniformizable Cauchy spaces (and Cauchy continuous maps) is a bireflective subconstruct of **Chy** and a bicoreflective subconstruct of **Unif** (cf. [11;1.3.]). Obviously, *the uniformizable Cauchy spaces are exactly the subspaces in* **Fil** *of all completely uniformizable topological spaces.*

3.16 Proposition. UChy *is a bireflective subconstruct of* **SubTop**.

Proof. Since **Chy** is bireflective in **Fil**, Proposition 3.16. follows immediately from Remark 3.15. ∎

3.17 Remark. Frič and Kent [7] studied completely regular Cauchy spaces as a generalization of uniformizable Cauchy spaces. Their definition coincides with the definition of completely regular filter spaces used by Bentley and Lowen–Colebunders [4], where a filter space is called *completely regular* provided that it is a subspace (in **Fil**) of some completely regular topological space (considered as filter space). A filter space (X, γ) is completely regular iff for each $\mathcal{F} \in \gamma$ the subfilter $\mathcal{G} = \{G \subset X : F$ is completely within G for some $F \in \mathcal{F}\}$ belongs to γ; here F is completely within G provided that there is a Cauchy continuous map $f : (X, \gamma) \to ([0,1], \gamma_t)$ such that $f[F] \subset \{0\}$ and $f[X \backslash G] \subset \{1\}$, where γ_t denotes the set of all convergent filters on the unit interval $[0,1]$ endowed with the usual topology (cf. [4]). The construct **CRegFil** of completely regular filter spaces (and Cauchy continuous maps) is bireflective in **Fil** (cf. [4;3.2.]). The restriction of the bicoreflector $\mathcal{C} : \mathbf{Fil} \to \mathbf{KConv_s}$ leads to the result that the construct **CRegTop** of completely regular topological spaces (and continuous maps) is bicoreflectively embedded into **CRegFil**. **CRegFil** is related to other important constructs mentioned in this paper by means of the following diagram, where r (resp. c) stands for embedding as a bireflective (resp. bicoreflective) subconstruct:

$$\begin{tikzcd}
 & \mathbf{SUConv} & \mathbf{Fil} \arrow[l, "c", "r"'] & \mathbf{KConv_s} \arrow[l, "c"'] \\
 & & \mathbf{SubLim} \arrow[u, "r"'] & \mathbf{Lim_s} \arrow[l, "c"'] \arrow[u, "r"'] \\
 & & \mathbf{SubPrTop} \arrow[u, "r"'] & \mathbf{PrTop_s} \arrow[l, "c"'] \arrow[u, "r"'] \\
 & & \mathbf{SubTop} \arrow[u, "r"'] & \mathbf{Top_s} \arrow[l, "c"'] \arrow[u, "r"'] \\
\mathbf{ULim} \arrow[uuuur, "r"] & \mathbf{Chy} \arrow[l, "c"'] \arrow[uuur, "r"] & & \\
\mathbf{Unif} \arrow[u, "r"] & & \mathbf{CRegFil} \arrow[ul, "r"] \arrow[uu, "r"'] & \mathbf{CRegTop} \arrow[l, "c"'] \arrow[uu, "r"'] \\
 & & \mathbf{UChy} \arrow[u, "r"'] \arrow[ull, "c"] &
\end{tikzcd}$$

4 Problem

Find a nice characterization of subspaces in **Fil** of symmetric pseudotopological spaces (considered as filter spaces) where a generalized convergence space (X, q) is called pseudotopological provided that $(\mathcal{F}, x) \in q$ whenever $(\mathcal{U}, x) \in q$ for each ultrafilter $\mathcal{U} \supset \mathcal{F}$.

References

[1] J. Adámek, J., H. Herrlich and G.E. Strecker. *Abstract and Concrete Categories*, Wiley–Intersci. Publ., New York, 1990.

[2] H.L. Bentley. Nearness spaces and extension of topological spaces, in *Studies in Topology*, Academic Press, New York, 1975, 47–66.

[3] H.L. Bentley, H. Herrlich and E. Lowen–Colebunders. Convergence, *J. Pure and Applied Algebra*, **68** (1990), 27–40.

[4] H.L. Bentley and E. Lowen–Colebunders. The cartesian closed topological hull of the category of completely regular filter spaces, *Cahiers Topol. Géom. Diff.*, **23**–4 (1992), 345–360.

[5] E. Čech. *Topological Spaces*, revised edition by M. Frolik and M. Katětov, Interscience, London, 1966.

[6] H.R. Fischer. Limesräume, *Math. Ann.*, **137** (1959), 269–303.

[7] R. Frič and D.C. Kent. On the natural completion functor for Cauchy spaces, *Bull. Austral. Math. Soc.*, **18** (1978), 335–343.

[8] H. Herrlich. Topological structures, in *Math. Centre Tracts*, Math. Centrum, Amsterdam, **52** (1974), 59–122.

[9] M. Katětov. On continuity structures and spaces of mappings, *Comm. Math. Univ. Carolinae*, **6** (1965), 257–278.

[10] H.–J. Kowalsky. Limesräume und Komplettierung, *Math. Nachrichten*, **12** (1954), 301–340.

[11] 11 E. Lowen–Colebunders. *Function Classes of Cauchy Continuous Maps*, Marcel Dekker, New York, 1989.

[12] G. Preuß. *Theory of Topological Structures*, Reidel, Dordrecht, 1988.

[13] G. Preuß. Improvement of Cauchy spaces, Q & A, *in General Topology*, **9** (1991), 159–166.

[14] G. Preuß. 1993, Cauchy spaces and generalizations, *Math. Japonica*, **38** (1993), 803–812.

[15] G. Preuß. Completion of semiuniform convergence spaces, in *Categorical Methods in Algebra and Topology*, H.–E. Porst, editor, Mathematik–Arbeitspapiere, Universität Bremen, **48** 1997, 353–361.

[16] G. Preuß. Convenient Topology, *Math. Japonica*, **47** (1998), 171–183.

[17] A. Weil. *Sur les espaces à structures uniformes et sur la topologie générale*, Hermann, Paris, 1937.

G. Preuß
Department of Mathematics
Free University of Berlin, Germany
`preuss@math.fu-berlin.de`

The Naturals are Lindelöf iff Ascoli Holds

Y.T. Rhineghost

ABSTRACT It is shown that in **ZF** (i.e., Zermelo–Fraenkel set theory without the Axiom of Choice) the classical Ascoli Theorem holds iff $\mathbb{N}$ is a Lindelöf space.

Key words: Compact, Lindelöf, function space, axiom of (countable) choice.

AMS Subject Classifications (2000): 03E25, 54D20, 54D30, 54C35.

0 Introduction

In [5] the authors asked "When is $\mathbb{N}$ Lindelöf?" and they answered this question by exhibiting several equivalent topological, resp. set theoretic, conditions. In particular they demonstrated that in **ZF**, $\mathbb{N}$ is Lindelöf iff the Axiom of Countable Choice holds for subsets of $\mathbb{R}$. It is known that the latter condition need not hold in **ZF** (see, e.g., [6]), and that it is properly weaker than the Axiom of Countable Choice (see [10]). Here it will be shown that another equivalent condition is the Ascoli Theorem in its classical form. Recall that the Ascoli Theorem appears in many variants (see, e.g., [9]), and that in a more modern and familiar form it is equivalent to the Boolean Prime Ideal Theorem (see [3]), provided that compactness is defined via the familiar Heine–Borel covering property, — resp., to the Axiom of Choice (see [4]), provided that compactness is defined via the Alexandroff–Urysohn property stating that each infinite set has a complete accumulation point. In this note a further variant of the Classical Ascoli Theorem will be exhibited, and it will be shown first that this variant holds in **ZF**.

I would like to thank Horst Herrlich and Evangeliste Jimplecute for many enlightening discussions on the topic of this note.

1 The Ascoli Theorem

In this section terminology will be fixed (in **ZF**).

Definitions 1.1

(1) A *function* is a continuous map $f : \mathbb{R} \longrightarrow \mathbb{R}$.

(2) A sequence (f_n) of functions *converges continuously* to a map f provided that the following holds:

$$\forall x \in \mathbb{R}\ \forall (x_n) \in \mathbb{R}^{\mathbb{N}}\ \ ((x_n) \longrightarrow x \Longrightarrow (f_n(x_n)) \longrightarrow f(x)).$$

Remarks 1.2

(1) If a sequence of functions converges continuously to a map f, then f is a function.

(2) If (f_n) converges locally uniformly to f, then (f_n) converges continuously to f, and if (f_n) converges continuously to f, then (f_n) converges pointwise to f.

Theorem 1.3 (Classical Ascoli Theorem) *For sets F of functions, the following conditions are equivalent:*

(1) Each sequence in F has a subsequence that converges continuously to some function g (not necessarily in F).

(2) (a) For each $x \in \mathbb{R}$, the set $F(x) = \{f(x) \mid f \in F\}$ is bounded, and

(b) F is equicontinuous, i.e.,

$$\forall x \in \mathbb{R}\ \forall \varepsilon > 0\ \exists \delta > 0\ \forall f \in F\ \forall y \in \mathbb{R}\big(|x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon\big).$$

Theorem 1.4 (Modified Ascoli Theorem) *For sets F of functions the following conditions are equivalent:*

(1) Each sequence in F has a subsequence that converges continuously to some function g (not necessarily in F).

(2) (a) For each countable subset G of F and each $x \in \mathbb{R}$ the set $G(x)$ is bounded, and

(b) each countable subset of F is equicontinuous.

2 Results

Proposition 2.1 *The Modified Ascoli Theorem holds.*

Proof. (1) $\Rightarrow$ (2a) Let G be a countable subset of F, say $G = \{f_n \mid n \in \mathbb{N}\}$ for some sequence (f_n) in F. Assume that $G(x)$ is unbounded for some $x \in \mathbb{R}$. For each $n \in \mathbb{N}$ define $\nu(n) = \min\{m \in \mathbb{N} \mid n < |f_m(x)|\}$. Then the sequence $(f_{\nu(n)})$ is such that no subsequence of $\big((f_{\nu(n)}(x)\big)$ converges, violating (in view of Remark 1.2 (2)) condition (1).

(1) $\Rightarrow$ (2b) Let $G = \{f_n | n \in \mathbb{N}\}$i be as above and assume that G is not equicontinuous at some $x \in \mathbb{R}$. Then there exists $\varepsilon > 0$ such that for each $\delta > 0$ there exist $n \in \mathbb{N}$ and $y \in \mathbb{R}$ with $|x - y| < \delta$ and $|f_n(x) - f_n(y)| \geq \varepsilon$.

For each $n \in \mathbb{N}$ define $\nu(n), g_n$ and x_n as follows:
$\nu(n) = \min\{m \in \mathbb{N} \exists y \in [x - 2^{-n}, x + 2^{-n}] \ \ |f_m(x) - f_m(y)| \geq \varepsilon\}$,
$g_n = f_{\nu(n)}$,
$x_n = \min\{y \in [x - 2^{-n}, x + 2^{-n}] \mid |g_n(x) - g_n(y)| \geq \varepsilon\}$.

Then $(x_n) \longrightarrow x$ and $|g_n(x) - g_n(x_n)| \geq \varepsilon$ for each $n \in \mathbb{N}$. Consequently no subsequence of (g_n) converges continuously at x, violating condition (1).

(2) $\Rightarrow$ (1) Express the rationals as a sequence (r_n), and let (f_n) be a sequence in F. By induction define a sequence of pairs (a_n, s_n) with $a_n \in \mathbb{R}$ and $s_n = (g^n_m)_{m \in \mathbb{N}}$ a sequence in F as follows:

1. Let a_0 be the smallest accumulation point of the sequence $((f_n(r_0))$. Define $s_0 = (g^0_n)_{n \in \mathbb{N}}$ by induction as a subsequence $(f_{\nu(n)})$ of (f_n) as follows:

$$\begin{array}{lrcl} \text{a)} & \nu(0) & = & \min\{m \in \mathbb{N} \mid |f_m = (r_0) - a_0| < 1\}, \\ \text{b)} & \nu(n+1) & = & \min\{m \in \mathbb{N} \mid \nu(n) < m \text{ and } |f_m(r_0 - a_0)| < \frac{1}{n+1}\}. \end{array}$$

Then $s_0 = (g^0_n) = (f_{\nu(n)})$ is a subsequence of (f_n) and $(g^0_n) \longrightarrow a_0$.

2. Let a_{n+1} be the smallest accumulation point of the sequence $s_n(r_{n+1}) = (g^n_m(r_{n+=1}))_{m \in \mathbb{N}}$. Define $s_{n+1} = (g^{n+1}_m)_{m \in \mathbb{N}}$ by induction as a subsequence $(g^n_{\nu(m)})$ of the sequence $s_n = (g^n_m)$ as follows:

$$\begin{array}{lrcl} \text{a)} & \nu(0) & = & \min\{p \in \mathbb{N} \mid |g^n_p(r_{n+1}) - a_{n+1}| < 1\}, \\ \text{b)} & \nu(m+1) & = & \min\{p \in \mathbb{N} \mid \nu(m) < p \text{ and } |g^n_p(r_{n+1}) - a_{n+1}| < \frac{1}{m+1}\}. \end{array}$$

Then the following hold:

(α) s_0 is a subsequence of (f_n).
(β) For each $n \in \mathbb{N}$, s_{n+1} is a subsequence of s_n.
(γ) For each $n \in \mathbb{N}$, the sequence $s_n(r_n) = (g^n_m(r_n))_{m \in \mathbb{N}}$ converges to a_n.

Next, consider the diagonal–sequence $s = (g^n_n)$. Then:

(α') s is a subsequence of (f_n).
(β') For each $n \in \mathbb{N}$, s is cofinal with a subsequence of s_n.

Thus

(γ') For each $n \in \mathbb{N}$, the sequence $s(r_n) = (g^m_m(r_n))_{n \in \mathbb{N}}$ converges to a_n.

So the subsequence $s = (g^n_n)_{n \in \mathbb{N}}$ of (f_n) has the property that $s(x) = (g^n_n(x))$ converges for each $x \in \mathbb{Q}$. Since $\mathbb{Q}$ is dense in $\mathbb{R}, G = \{g^n_n | n \in \mathbb{N}\}$ is equicontinuous and $\mathbb{R}$ is complete, the familiar arguments imply that s converges locally uniformly and thus continuously to some function a with $a(r_n) = a_n$ for each $n \in \mathbb{N}$. Consequently (1) holds. ∎

Theorem 2.2 *Equivalent are:*
(a) The Classical Ascoli Theorem,
(b) $\mathbb{N}$ *is a Lindelöf space.*

Proof. In [5] it is shown that (b) is equivalent to the condition:
(c) Every unbounded subset of $\mathbb{R}$ contains a countable unbounded subset.

Moreover, since the cardinality of the set of functions agrees with the cardinality of $\mathbb{R}$, condition (b) is equivalent to
(d) the Axiom of Countable Choice for sets of functions.

Thus (see [5]) it remains to show the implications (a) $\Rightarrow$ (c) and (d) $\Rightarrow$ (a).

(a) $\Rightarrow$ (c) Let B be an unbounded subset of $\mathbb{R}$. Consider, for each $b \in B$ the constant function $f_b\colon \mathbb{R} \longrightarrow \mathbb{R}$ with value b. Then the set $F = \{f_b \mid b \in B\}$ violates condition (2a) of the Classical Ascoli Theorem, thus — by (a) — also condition (1). The latter is identical with condition (1) of the Modified Ascoli Theorem. In view of Proposition 2.1 and the fact that F is equicontinuous, this implies that there is a countable subset G of F and an $x \in \mathbb{R}$ such that the set $G(x)$ is unbounded. Thus $G(x)$ is a countable unbounded subset of B.

(d) $\Rightarrow$ (a) By Proposition 2.1, (2) implies (1) in the Classical Ascoli Theorem. Thus it remains to be shown that in the presence of (d), (1) implies (2). Let F satisfy (1).

(α) Consider $x \in \mathbb{R}$. If $F(x)$ is unbounded, then $F_n = \{f \in F \mid |f(x)| > n\} \neq \emptyset$ for each n. Thus (d) implies that there exists an element (f_n) of ΠF_n. Hence $G = \{f_n \mid n \in \mathbb{N}\}$ is a countable subset of F with $G(x)$ being unbounded. This contradicts Proposition 2.1.

(β) Assume that F fails to satisfy condition (2b). Then there exist $x \in \mathbb{R}$ and $\varepsilon > 0$ such that for each $n \in \mathbb{N}$ the set $F_n = \{f \in F \mid \exists y \in \mathbb{R} \ \ |x-y| < 2^{-n}$ and $|f(x) - f(y)| \geq \varepsilon\}$ is non-empty. Thus (d) implies that there exists an element (f_n) of ΠF_n. Consequently $G = \{f_n \mid n \in \mathbb{N}\}$ is a countable subset of F that fails to be equicontinuous. This contradicts Proposition 2.1. ∎

3 Remarks on Games

The Classical Ascoli Theorem, being equivalent to the countable axiom of choice for subsets of the reals, must be considered as a weak form of the axiom of choice. Thus it is somewhat surprising that it also follows from the axiom of determinateness (see [7, 8]), since the latter implies that $\mathbb{R}$ cannot be well-ordered ([6, Theorem 12.14]), hence that the axiom of choice does not hold for subsets of $\mathbb{R}$.

In fact, it is easy to see that the countable axiom of choice for subsets of the reals is equivalent to a weak form of the axiom of determinateness:

Consider the game $G(\mathbf{N}^{\mathbb{N}}, A)$ described as follows: Two players chose alternately natural numbers $x_0, x_1, x_2, \ldots$, each of the players knowing at each move all the previously chosen elements. The first player wins if the resulting sequence (x_n) belongs to A, otherwise the second player wins. The axiom **A** of determinateness can be stated as follows:

A: For each subset A of $\mathbb{N}^{\mathbb{N}}$ one of the players has a winning strategy for the game $G(\mathbb{N}^{\mathbb{N}}, A)$.

Call a subset A of $\mathbb{N}^{\mathbb{N}}$ *saturated* provided that it contains with each element (a_n) all elements (x_n) of $\mathbb{N}^{\mathbb{N}}$ that satisfy $x_n = a_n$ for $n = 0$ and for all odd n. Then the axiom of countable choice for subsets of the reals (and thus the classical Ascoli Theorem) is equivalent to the following weak form of **A**:

$\mathbf{A_W}$: For each saturated subset A of $\mathbb{N}^{\mathbb{N}}$ one of the players has a winning strategy for the game $G(\mathbb{N}^{\mathbb{N}}, A)$.

This can be seen easily via the game $G(\mathbb{N} \times R, A)$, described as follows: The first player chooses a natural number n, then the second player — knowing n — chooses a real number r. The second player wins if $(n, r) \in A$, otherwise the first player wins. Trivially the axiom of countable choice for subsets of the reals holds if and only if the following is true:

$\tilde{\mathbf{A}}$: For each subset A of $\mathbb{N} \times R$ one of the players has a winning strategy for the game $G(\mathbb{N} \times R, A)$.

That $\mathbf{A}_W$ and $\tilde{\mathbf{A}}$ are equivalent follows immediately from the existence of a bijection $R \longrightarrow \mathbb{N}^{\mathbb{N}}$ (compare [7]).

References

[1] C. Arzelà. Funzioni di linee, *Atti della Reale Accademia dei Lincei, Rendiconti*, **5** (1889), 342–348.

[2] G. Ascoli. Le Curve limite di una varietà data di curve, *Mem. Accad. Lincei*, (3) **18** (1887), 521–586.

[3] H. Herrlich. The Ascoli Theorem is equivalent to the Boolean Prime Ideal Theorem, *Rostock Math. Kolloq.*, **51** (1997), 137–140.

[4] H. Herrlich. The Ascoli Theorem is equivalent to the Axiom of Choice. *Seminarbericht, FB Mathematik, Univ. Hagen*, **62** (1997), 97–100.

[5] H. Herrlich and G.E. Strecker. When is $\mathbb{N}$ Lindelöf?, *Comment. Math. Univ. Carolinae*, **38** (1997), 553–556.

[6] J.T. Jech, J.T., *The Axiom of Choice*, North Holland, Amsterdam, 1973.

[7] J. Mycielski. On the axiom of determinateness. *Fund. Math.*, **53** (1964), 205–224.

[8] J. Mycielski and H. Steinhaus. A mathematical axiom contradicting the axiom of choice, *Bull. Acad. Polon. Sci.*, **10** (1962), 1–3.

[9] H. Poppe. *Compactness in General Function Spaces*, Deutscher Verl. Wiss., Berlin, 1974.

[10] G. Schlitt. The Lindelöf Tychonoff Theorem and choice principles, *Math. Proc. Cambr. Phil. Soc.*, **110** (1991), 57–65.

Y.T. Rhineghost
c/o H. Herrlich
Department of Mathematics
University of Bremen
P.O. Box 33 o4 4o
28334 Bremen, Germany
herrlich@math.Uni-Bremen.de

a homeomorphism, then we have a group action of $\mathbb{Z}$ on X . For a linear system (E, L) the action generated by L is linear so that E regarded as an Abelian group becomes a $\mathbb{Z}^+$-module. In [17] W. Parry and S. Tuncel study the group action of $\mathbb{Z}$ on ℓ_2 generated by a unitary operator $L : \ell_2 \to \ell_2$. Observing that the expression $0x+1x+\cdots+(n-1)x$ regarded as a mapping from $\mathbb{Z}$ to ℓ_2 is a one-cocycle for every vector $x \in \ell_2$, they call every element $x \in \ell_2$ a one-cocycle and if x can be presented in the form $x = y - Ty$ with $y \in \ell_2$, then it is called a coboundary. They show that the one-cohomology of this action has a dynamical meaning by proving that the norm $\|x+Tx+\cdots+T^{n-1}x\|$ of the vector $x+Tx+\cdots+T^{n-1}x$ is bounded as $n \to \infty$ if and only if x is a coboundary. In Section 2 we carry out a similar investigation using another $\mathbb{Z}^+$-module, namely the set $S(E)$ defined as the set of all functions $f : E \to \mathbb{R}$ which can be presented in the form $\sum_{k=1}^{n} a_k\|L_k x\|_k$ where $a_k \in \mathbb{R}, L_k \in \mathcal{L}(E)$, $\|\cdot\|_k \in N(E)$, $k = 1, 2, \cdots, n$, $n \in \mathbb{N}$ and where $\mathcal{L}(E)$ is the Banach algebra of all linear endomorphisms of E and $N(E)$ the set of all compatible norms $\|\cdot\|$ on E. It is evident that $S(E)$ is L-invariant since if $f \in S(E)$, then $f(Lx) = \sum_{k=1}^{n} a_k\|L_k Lx\|_k$ which again belongs to $S(E)$. Thus $S(E)$ is a $\mathbb{Z}^+$-module which we shall use when studying properties of L relative to various equivalent norms on E. Since, as it follows immediately from the definition, the functions of $S(E)$ are bounded on bounded sets of E, the module $S(E)$ is suitable for studying the summability of series of the form $\sum_{n=0}^{\infty}\|L^n x\|$, $x \in E$ and $\|\cdot\| \in N(E)$. The main attention in Section 2 will be given to systems (X, T) which are of contraction type, i.e., where T is a Banach contraction relative to a metric d on X which is compatible with its topology. Adjusting the ideas of C. Moore and K. Schmidt [14] to our situation, we develop in Section 3 a method based on cohomology for obtaining invariants of a system (X, T). Finally in the last section we introduce the notion of "semilinearization" of a system (X, T). It turns out that there are systems (X, T) important in analysis, for which X is finite dimensional but $D(X, T) = \infty$. We ask whether such a system can be topologically embedded into a Euclidean space E^n so that $A \circ i(x) \le i(Tx)$ where A is an $n \times n-$matrix with nonnegative entries and $\le$ is the usual partial order relation between vectors in E^n. If this is possible we say that (X, T) can be semilinearized in (E^n, A). We use this method to obtain a bothsided estimate of iterated sine function $\sin^{(r)} x$ in the form $x - \frac{r}{6}x^3 \le \sin^{(r)} x \le x - \frac{r}{6}x^3 + \frac{r}{120}(5r-4)x^5$, valid on the interval $[0, 3\sqrt{2}]$ for $r \in \mathbb{N}$.

2 Some old and new results

In 1959 J. de Groot ([2], [6]) produced a universal model (ℓ_2, L) of a linear system in which every system (X, T) can be linearized. Later, in 1961 with A.H. Copeland ([3]), they found a similar universal model (ℓ_2, L') with L' invertible for all systems (X, T) where T is an autohomeomorphism of X. The exact statements are as follows:

Theorem A. *There exists a linear system (ℓ_2, L) in which every system (X, T) can be linearized. Moreover the embedding $i : X \to \ell_2$ is such that the closure in ℓ_2 of the image $i(X)$ is compact.*

Theorem B. *There exists a linear system (ℓ_2, L') with L' invertible in which every system (X, T) with T being an autohomeomorphism can be linearized. Moreover, the embedding $i : X \to \ell_2$ is such that the image $i(X)$ has compact closure in ℓ_2.*

The idea of the proof (see [3]) is based on interpreting ℓ_2 as the set of double indexed sequences $(x_{i,n})$ with $\sum_{i=1}^{\infty}\sum_{n=0}^{\infty} x_{i,n}^2 < \infty$ and identifying the space X as a subset of the Hilbert cube of the sub-Hilbert space consisting of sequences of the form $(x_{i,o})$. The universal linear operator L is defined by $(Lx)_{i,n} = 2x_{i,n+1}$ for $i = 1, 2, \ldots$ and $n = 0, 1, \ldots$. Modifying this construction by letting the second index n range through the set $\mathbb{Z}$ of all integers, they arrive at the proof of Theorem B where the universal invertible operator L' is defined by

$$(L'x)_{i,n} = \begin{array}{ll} 2x_{i,n+1} & \text{if} \quad n \geq 0 \ , \\ \frac{1}{2}x_{i,n+1} & \text{if} \quad n < 0 \ . \end{array}$$

Our main objective will be to linearize systems which represent a Banach contraction relative to a suitable metric and which we call B-systems.

Definition 2.1. A system (X, T) is called a B-system if T has a unique fixed point $x_0 \in X$ and there exists a metric $d \in M(X)$ relative to which T is a Banach contraction. Here and in the sequel $M(X)$ denotes the set of all compatible metrics on X.

The following statement gives a purely topological characterization of B-systems where no reference to any metric notion is involved.

Theorem C. *A system (X, T) is a B-system if and only if it has the following two properties:*

(i) T has a unique fixed point $x_0 \in X$ and for every $x \in X$ the orbit $\{T^n x : n \in \mathbb{N}\}$ converges to it.

(ii) There exists an open set U containing x_0 such that for every open set V containing x_0 there exists $n \in \mathbb{N}$ such that $m \geq n$ implies $T^m(U) \subseteq V$.

For the proof see [13] or [16].

For compact B-systems there is a simpler version of this characterization.

Theorem C′. *If (X,T) is a system with X compact, then the following two properties of (X,T) are equivalent:*

(A1) (X,T) is a B-system.

(A2) The intersection $\bigcap\{T^n X : n \in \mathbb{N}\}$ is a singleton.

For the proof see [7].

Using Theorem C we prove our main result.

Theorem 2.1. *A linear system (E,L) is a B-system if and only if there is a compatible norm $\|\cdot\|$, i.e., $\|\cdot\| in N(E)$, such that $\|L\| < 1$.*

Proof. The if part follows from the definition, since $\|L\| < 1$ says that L is a Banach contraction relative to the metric induced by the norm $\|\cdot\|$. To prove the converse assume (E,L) is a B-system and let $\|\cdot\| \in N(E)$. Expressing the property (ii) of Theorem C in terms of open balls $B(r)$ in E relative to the norm $\|\cdot\|$, we conclude that for every $a \in \mathbb{R}$ satisfying $0 < a < 1$ there exists some $n \in \mathbb{N}$ such that $m \geq n$ implies $\|L^m x\| \leq a\|x\|$ for every $x \in E$. Applying this inequality to $m = n$ and subsequently to $Lx, L^2x, \ldots,$ we obtain the inequalities

$$\begin{aligned} &\|L^{rn}x\| \leq a^r\|x\| \ , \\ &\|L^{rn+i}x\| \leq a^r\|L^i x\| \end{aligned} \tag{1}$$

for $r \in \mathbb{N}$ and $i = 1, 2, \ldots, n-1$. From (1) it follows that the series $\sum_{k=0}^{\infty}\|L^k x\|$ converges, but we need a little bit more. We must find a number $c > 1$ such that the series $\sum_{k=0}^{\infty} c^k\|L^k x\|$ still converges. Consulting (1) we see that this series is majorized by the series $\|x\|(1 + c\|L\| + \cdots + c^{n-1}\|L^{n-1}\|)\sum_{k=0}^{\infty} a^k c^{kn}$ so that this series converges if we choose c so that $ac^n < 1$, i.e., $c < a^{-\frac{1}{n}}$. The series defines a compatible norm, say $|||\cdot||| \in N(E)$ and for the corresponding norm $|||L|||$ of L we obtain

$$|||L||| = \sup \frac{\sum_{k=0}^{\infty} c^k\|L^{k+1}x\|}{\sum_{k=0}^{\infty} c^k\|L^k x\|} \leq \frac{1}{c} = a^{\frac{1}{n}} < 1$$

which was to be proved. ■

Remark 2.1. We use the letter L generically to denote linear continuous operators. It should not be confused with the special operators L and L' appearing in Theorems A and B.

Remark 2.2. According to the definition of a B-system, a linear B-system (E, L) is such that for some $a \in (0,1)$ one can find a metric $d \in M(E)$ relative to which L is a Banach contraction with the Lipschitz constant a. The question remains whether this metric d can be found in the form $d(x,y) = \|x - y\|$ where $\|\cdot\| \in N(E)$. Theorem 2.1. says that it is so.

Remark 2.3. The method of linearization $i : (X,T) \to (E,L)$ of a system (X,T) described in Theorems A and B is characterized by the fact that the norms of the linear operators L involved are > 1 and that the image $i(X)$ has compact closure $\overline{i(X)}$ in E. This implies that the system $(\overline{i(X)}, L^*)$, where L^* is the restriction of L to $\overline{i(X)}$, is a compactification of the original system (X,T) in the sense that (X,T) is isomorphic to a dense subsystem of $(\overline{i(X)}, L^*)$. From this follows that if (X,T) is a B-system such that some of its compactifications is no longer a B-system, then we cannot expect that the linear system (E,L) is a B-system. This is a shortcoming of this method which will be referred to as the CG-method in the sequel. However we would like to make it clear that so far as we know no other systematic method has been developed for general noncompact systems. Next we give two examples where a non CG-linearization can easily be constructed.

Example 2.1. Let (X,T) be a system with $X = [0,1)$ and T defined by the quadratic function $Tx = x^2$. We observe that (X,T) is a B-system since both the conditions (i) and (ii) of Theorem C are satisfied. But since (X,T) has a compactification on $[0,1]$ which is no longer a B-system (the extension of T on $[0,1]$ has two fixed points), the CG-method would furnish a linearization (E,L) which may not be a B-system. On the other hand there exists a simple linearization in $(E^1, \frac{1}{2}I)$ where I is the identity operator, obtained by the map $i : [0,1) \longrightarrow \mathbb{R}$ defined by $i(x) = (\log x)^{-1}$ if $x \neq 0$ and $i(0) = 0$.

Example 2.2. The system (X,T) is the same as above but the linearization will be in ℓ_2, obtained by a suitable modification of the CG-method so that the resulting linear operator will have norm < 1. If $x \in [0,1)$, then its T-orbit is $(x, x,^2, \dots, x^{2^n}, \dots)$ which is a point in ℓ_2 as it is readily seen. If c is an arbitrary constant > 1, then the sequence $(x, cx^2, \dots, c^n x^{2^n}, \dots)$ is still an element of ℓ_2, so we can define the embedding $i : [0,1) \to \ell_2$ setting $i(x) = (x, cx^2, \dots, c^n x^{2^n}, \dots) \in \ell_2$. We observe that $i(Tx) = i(x^2) = (x^2, cx^4, \dots, c^n x^{2^{n+1}}, \dots)$ so that we have $i \circ T = A \circ i$ where $A : \ell_2 \to \ell_2$ is defined as the composition of the shift $(x_0, x_1, \dots, x_n, \dots) \to (x_1, x_2, \dots, x_{n+1}, \dots)$ followed by multiplication by c^{-1}. Now we must show that $i : [0,1) \to \ell_2$ is a topological embedding. We cannot use the compactness argument as in [3] since the closure of $i[[0,1)]$ is not compact (the norm of $(x, cx^2, \dots, c^n x^{2^n}, \dots)$ increases to ∞ as $x \to 1$).

We proceed as follows: We contemplate the function $\varphi(x) = \sum_{k=0}^{\infty} c^k x^{2^k}$ (this function is known to be analytic in the unit disc of the complex plane) and observe that the norm $\|i(y) - i(x)\|$ for $x, y \in [0, 1)$ is majorized by the expression $|y - x|\varphi'(y)$. Since the derivative $\varphi'(y)$ is continuous the inequality $|y - x|\varphi'(y) \leq \epsilon$, where ϵ is any number > 0, has a solution furnishing the needed $\vartheta > 0$ to prove the continuity of i at x. The inverse i^{-1} is continuous since it can be considered as the orthogonal projection of $i(x)$ to its 0-th coordinate x. The norm $\|A\|$ of the linearizing operator A is $c^{-1} < 1$, and since $c > 1$ is arbitrary we can produce linearization of this particular system (X, T) with arbitrary small norm. This leads us to ask the following.

Question 2.1. Suppose that a B-system (X, T) has the property that for every $\epsilon > 0$ there exists a linearization (E, L) of it with $\|L\| \leq \epsilon$. Does this imply the existence of a linearization (E, L) in which the spectral radius of L is zero? This means is $\lim\limits_{n\to\infty} \|L^n\|^{\frac{1}{n}} = 0$?

This conjecture is supported by the following example.

Example 2.3. Let $C[0, 1]$ be the Banach space of continuous real valued functions on $[0, 1]$ equipped with supremum norm and let $L : C[0, 1] \to C[0, 1]$ be the operator of integration defined by $(Lf)(x) = \int_0^x f(t)dt$ for $f \in C[0, 1]$. By iterating L we obtain $(L^n f)(x) = \int_0^x \frac{(x-t)^{n-1}}{(n-1)!} f(t)dt$ from which it follows easily that $\|L^n\| \leq (n!)^{-1}$ so that $\|L^n\|^{\frac{1}{n}} \leq (n!)^{-\frac{1}{n}}$. Since $(n!)^{\frac{1}{n}} \to \infty$ as $n \to \infty$, this implies that L is quasinilpotent, i.e., that its spectral radius is zero. For the sake of completeness we present now two results concerning linearization of compact systems.

Theorem D. *If (X, T) is a compact B-system, then given any constant $a \in (0, 1)$ there is a linearization (E, L) of (X, T) with $\|L\| \leq a$.*

For the proof see [4].

Theorem E. *If (X, T) is a compact system, not necessarily a B-system but such that the family of iterations $\{T^n : n \in \mathbb{N}\}$ is evenly continuous, then for every constant $a \in (0, 1)$ there exists a linearization of (X, T) in the product $(E_1, L_1) \times (E_2, L_2)$ where $\|L_1\| \leq a$ and L_2 is orthogonal on E_2.*

For the proof see [8].

These results have been exploited and generalized by us and by several other researchers to obtain analogous statements concerning flows, semi-flows and more general dynamical systems ([1],[4],[5],[8],[9],[11],[12]).

From Theorem C′ it follows that the characteristic property of a compact B-system (X, T) is that the intersection $\bigcap\{T^n X : n \in \mathbb{N}\}$ is a singleton. But it should be noted that there are non-compact B-systems with this property, and which is a little bit surprising, they may have a compactification which is no longer a B-system. We obtain an example of this situation by modifying Example 1, taking for X the rationals Q on $[0, 1)$

and for T the restriction to Q. An easy number theoretical argument shows that $\bigcap\{T^nQ : n \in \mathbb{N}\} = \{0\}$, whereas the compactification of (Q, T) is the system $([0,1], T^*)$ where the extension T^* has two fixed points, 0 and 1, so that $([0,1], T^*)$ is not a B-system.

There is still another question to clarify. Confronting theorems A and D one may think that if the system (X, T) is compact, then the linearization given by Theorem D can be obtained also by the CG-method, i.e., by Theorem A, if we reduce the linearization space (ℓ_2, L) to a subspace (E, L^*) generated by the image $i(X)$. The following examples shows it may not be so.

Example 2.4. Let $H \subseteq \ell_2$ be the Hilbert cube in ℓ_2, i.e., $H = \{(x_n) : 0 \le x_n \le \frac{1}{n}, n \in \mathbb{N}\}$ and let $\{a_n\}$ be an increasing sequence of positive numbers with $\lim a_n = 1$ as $n \to \infty$. Let $L : \ell_2 \to \ell_2$ be defined by $L(x_n) = (a_n x_n)$. From this follows that H is invariant under L so that the restriction L^* of L to H defines a system (H, L^*) and a simple argument shows that $\bigcap\{L^{*n}H : n \in \mathbb{N}\} = \{0\}$. Thus Theorem C$'$ implies that (H, L^*) is a B-system. On the other hand (ℓ_2, L) is not a B-system since it does not satisfy the condition (ii) of Theorem C. Indeed, this condition would in our case imply the existence of a number $n \in \mathbb{N}$ such that for every $x \in \ell_2$ with $\|x\| = 1$ we have $\|L^n x\| \le \frac{1}{2}$. If we choose for x the unit vector $e_k = (\underbrace{0, 0, \ldots, 1}_{k \text{ places}}, 00, \ldots)$ we obtain that $\|L^n e_k\| = a_k^n \le \frac{1}{2}$ which is impossible since $a_k \to 1$ as $k \to \infty$.

We shall now focus our attention on general linear systems (E, L) and study the action of the semigroup $\mathbb{Z}^+$ on the Abelian group $S(E)$ generated by L, i.e., $S(E)$ becomes a $\mathbb{Z}^+$-module induced by L. As mentioned in the introduction we shall show a certain relevance of one-cohomology of this module to dynamical properties of the operator L. Since we use only one-cohomology we shall omit in the sequel the reference to dimension and speak about cocycle and coboundaries instead of one-cocycles, etc. Elements $f \in S(E)$ are called cocycles; the action of L on f, $Lf \in S(E)$ is defined by $(Lf)(x) = f(Lx)$ for $x \in E$. An element $f \in S(E)$ is called a coboundary if there exists an element $g \in S(E)$ with $f = g - Lg$.

Definition 2.2. The linear system (E, L) is said to have bounded orbits if for every $x \in E$ the L-orbit $\{L^n x : n \in \mathbb{N}\}$ is a bounded subset of E. If L is invertible we say (E, L) has bothsided orbits bounded if $\{L^n x : n \in \mathbb{Z}\}$ is bounded for every $x \in E$.

Theorem 2.2. *If (E, L) has bounded orbits and there exists a norm $\|\cdot\| \in N(E)$ which is a coboundary, then there exists a norm $|||\cdot||| \in N(E)$ in which $|||L||| \le 1$.*

Proof. Since $\|\cdot\|$ is a coboundary there is a function $f \in S(E)$ such that $\|x\| = f(x) - f(Lx)$ so that

$$\sum_{k=0}^{n-1} \|L^k x\| = f(x) - f(L^n x) \quad . \tag{1}$$

Since the orbits are bounded and f is bounded on bounded sets, the left-hand side of (1) is bounded, which implies that the infinite sum $\sum_{k=0}^{\infty} \|L^k x\|$ converges and which in turn implies that $\|L^k x\| \to 0$ as $k \to 0$. This means that the orbit $\{L^k x : n \in \mathbb{N}\}$ converges to 0 and since $f(0) = 0$ we obtain that

$$\sum_{k=0}^{\infty} \|L^k x\| = f(x) \quad . \tag{2}$$

Observing that the left-hand side of (2) satisfies the algebraic properties of a norm, the equality (2) shows that this norm is continuous so that it belongs to $N(E)$. Denoting it by $|||\cdot|||$ we obtain for the corresponding $|||L|||$ that it equals

$$\sup_{x \neq 0} \frac{\sum_{k=0}^{\infty} \|L^{k+1} x\|}{\sum_{k=0}^{\infty} \|L^k x\|} = \sup_{x \neq 0} \frac{\sum_{k=1}^{\infty} \|L^k x\|}{\sum_{k=0}^{\infty} \|L^k x\|} \leq 1 \quad ,$$

which completes our proof. ■

There is a partial converse to this statement.

Theorem 2.3. *If (E, L) is such that in some norm $\|\cdot\| \in N(E)$ the corresponding norm $\|L\|$ is < 1, then the norm $\|\cdot\|$ is a coboundary.*

Proof. Since $\|L\| = a < 1$ we see that the series $\sum_{k=0}^{\infty} \|L^k x\|$ converges for every $x \in E$ since it is majorized by $\|x\| \sum_{k=0}^{\infty} a^k$. From this follows that the sum defines a compatible norm $|||\cdot||| \in N(E)$ and we have $\|x\| + |||Lx||| = |||x|||$ saying that $\|\cdot\|$ is a coboundary as claimed. ■

Remark 2.4. The norm $|||\cdot||| \in N(E)$ appearing in this proof satisfies the condition that $x \neq 0$ implies $|||Lx||| < |||x|||$. We may say that L is contractive in this norm but this contractivity is weaker than that in the sense of Banach contraction.

Definition 2.3. If (E, L) is a linear system, we say that a norm $\|\cdot\| \in N(E)$ is contractive for L if $x \neq 0$ implies $\|Lx\| < \|x\|$.

Definition 2.4. A linear system (E, L) is said to be contractive if there is a norm in $N(E)$ which is contractive for L and if for every $x \in E$ the orbit $\{L^n x\}$ converges to 0.

It turns out that also this type of contractivity has a nice purely topological formulation using the notion of an "attractor for compact sets" due to R. Nussbaum [15].

Theorem F. *If the linear system (E, L) is contractive, then the singleton $\{0\}$ is an attractor for compact sets, which means that for every compact set $K \subseteq E$ and every $r > 0$ there is $n \in \mathbb{N}$ such that $m \geq n$ implies that $L^m(K)$ is contained in the ball of radius r about* 0.

The proof follows from the main result of [15].

Theorem 2.4. *If the system (E, L) is contractive and the operator L is compact, then (E, L) is a B-system.*

Proof. Let $\|\cdot\| \in N(E)$ be the norm which is contractive for L. Then the unit ball $B \subseteq E$ relative to this norm is L-invariant and we have $\overline{L(B)} \subseteq B$ with $\overline{L(B)}$ compact. Applying to it the attractor property of Theorem F and choosing $r = \frac{1}{2}$, this theorem furnishes a number $n \in \mathbb{N}$ such that $m \geq n$ implies $\|L^m x\| \leq \frac{1}{2}$ for every $x \in \overline{L(B)}$ and therefore $\|L^{m+1}x\| \leq \frac{1}{2}$ for every $x \in B$. But from this follows that the condition (ii) of Theorem C is satisfied so that (E, L) is a B-system as stated. ■

Remark 2.5. From Theorem 2.1, it follows that the property of being a B-system implies that of being contractive. However the converse is not true in general unless the operator L is compact. The linear system (ℓ_2, L) of Example 2.4. has been shown not to be a B-system but it is contractive. Indeed, the natural norm $\|\cdot\|$ of ℓ_2 is contractive for L since $\|L(x_n)\|^2 = \sum_{n=0}^{\infty} a_n^2 x_n^2 < \sum_{n=0}^{\infty} x_n^2$ if $(x_n) \neq 0$; that $L^k(x_n) \to 0$ as $k \to \infty$ we prove as follows: For $(x_n) \in \ell_2$ and $\epsilon > 0$ let $n \in \mathbb{N}$ be such that $\sum_{n+1}^{\infty} x_i^2 \leq \epsilon$. Then we have that

$$\|L^k(x_n)\|^2 = \sum_{i=0}^{\infty} a_i^{2k} x_i^2 \leq \sum_{i=0}^{n} a_i^{2k} x_i^2 + \epsilon \ .$$

This implies that $\lim \|L^k(x_n)\|^2 \leq \epsilon$ as $k \to \infty$ and since ϵ is arbitrary our assertion follows.

Finally we shall pay attention to linear systems (E, L) with L invertible.

Theorem 2.5. *Let (E, L) be a linear system with L invertible and such that for every $x \in E$ the full orbit $\{L^n x : n \in \mathbb{Z}\}$ is bounded. Then there is no norm $\|\cdot\| \in N(E)$ which is a coboundary.*

Proof. Suppose there is such a norm $\|\cdot\|$. Then we have $\|x\| = f(x) - f(Lx)$ with $f \in S(E)$. Since the orbits are bounded this would imply that

$\sum_{n=0}^{\infty}\|L^n x\| = f(x)$ and also that $\sum_{n=1}^{\infty}\|L^{-n}x\| = -f(x)$ which is absurd since $f(x)$ must be > 0 for $x \neq 0$.

This simple result illustrates the dynamical meaning of the statement that a system has a norm which is a coboundary. ■

3 Dimension of Linearization and Invariants of Systems

For a system (X, T), we defined in Section 1 the number $D(X, T)$ as the minimum of those $n \in \mathbb{N}\bigcup\{\infty\}$ for which (X, T) can be linearized in E^n. It is an invariant of the system (X, T) since if $(X_i, T_i), i = 1, 2$ are isomorphic systems, i.e., $(X_1, T_1) \simeq (X_2, T_2)$, then $D(X_1, T_1) = D(X_2, T_2)$. It is also clear that if a system (X, T) can be embedded into a system (Y, S), then we have $D(X, T) \leq D(Y, S)$.

The most important question concerning $D(X, T)$ is to ask when it is finite, i.e., $D(X, T) < \infty$. For the system (X, T) in Example 2.1. we found that $D(X, T) = 1$. In [3] is proved that for a system (X, T) where $\dim X < \infty$ and where T is a periodic homeomorphism we have $D(X, T) < \infty$. So far as we know there is no criterion for the condition $D(X, T) < \infty$ applicable to systems (X, T) with $\dim X < \infty$ and arbitrary T.

Theorem 3.1. *If (X, T) is a system with T surjective but not one-to-one, then $D(X, T) = \infty$.*

Proof. By contradiction. Suppose $D(X, T) = n \in \mathbb{N}$ and (X, T) is linearized in (E^n, A) by the embedding $i : X \to E^n$. Since T is not one-to-one the matrix A must be singular so that $Ai(X)$ is contained in a linear subspace of dimension less than n. Due to the surjectivity of T, for any $x \in X$ there exists $y \in X$ with $x = Ty$. Applying the map i we have $i(x) = i(Ty) = Ai(y)$ so that $i(x) \in Ai(X)$. This is the desired contradiction since the dimension of AE is $< n$. ■

Example 3.1. Let $X = \{z : |z| = 1\}$ be the circle in the complex plane and $T : X \to X$ the quadratic function $Tz = z^2$. Since T is surjective but not one-to-one, Theorem 3.1 says that $D(X, T) = \infty$, which contrasts sharply with the similar system studied in Example 2.1.

Along with $D(X, T)$ one can construct other invariants based on linearization of (X, T), e.g., one may consider the minimum of all spectral radia ϱ of all linear operators L appearing in linearizations of (X, T). In their paper [3] Copeland and de Groot study autohomeomorphisms $T : X \to X$ of prime period p via the Lefschetz invariant of a map based on simplicial homology of finitely triangulable spaces. The most systematic method for constructing invariants of systems has been developed by C. Moore and K.

Schmidt in their paper [14] on ergodic theory. Their system consists of a locally compact topological group acting on a measure space. Modifying their construction and adapting it to our situation, i.e., considering the action of the semigroup $\mathbb{Z}^+$ on a separable topological space X induced by a continuous selfmap $T : X \to X$, we set up the following theory:

Definition 3.1. Given a space X and an Abelian topological group A let $D^n(X, A)$ denote the set of all continuous functions from $X^{n+1}((n+1)$-fold Cartesian product of X with itself) to A for $n \in \mathbb{Z}^+$, and let the map $d^n : D^n(X, A) \to D^{n+1}(X, A)$ be defined by $(d^n f)(x_0, x_1, \dots, x_{n+1}) = \sum_{i=0}^{n+1}(-1)^i f(x_0, \dots, \hat{x}_i, \dots, x_{n+1})$ for $n = 0, 1, 2, \dots$ and let $A \to D^\circ(X, A)$ be the map which sends $a \in A$ to the constant function $f(x) = a$ for every $x \in X$.

Proposition 3.1. *The sequence of groups and homomorphisms:*

$$0 \to A \to D^0(X, A) \xrightarrow{d^0} D^1(X, A) \to \cdots D^n(X, A) \xrightarrow{d^n} D^{n+1}(X, A) \to \cdots$$

is exact.

Proof. It is easy to verify that $d^{n+1} \circ d^n = 0$ for every $n \in \mathbb{Z}^+$ so that $Im(d^{n-1}) \subseteq Ker(d^n)$ and it remains to show that $Im(d^{n-1}) = Ker(d^n)$, which means that every cocycle is a coboundary. If $f \in Ker(d^n)$ then we have:

$$\sum_{i=1}^{n}(-1)^i f(x_0, \dots, \hat{x}_i, \dots, x_{n+1}) + (-1)^{n+1} f(x_0, \dots, x_n) = 0 \quad .$$

Choosing a fixed point $x \in X$ and defining the function $g : X^n \to A$ by setting $g(x_0, x_1, \dots, x_{n-1}) = f(x_0, x_1, \dots, x_{n-1}, x)$, we can rewrite the last equality as

$$\sum_{i=0}^{n}(-1)^i g(x_0, \dots, \hat{x}_i, \dots, x_n) = -(-1)^{n+1} f(x_0, \dots, x_n)$$

from which it follows that $\pm f$ is a coboundary $d^{n-1}g$, which was to be proved. ∎

Now comes into play the action of $\mathbb{Z}^+$ on X generated by T. We observe that T induces an action of $\mathbb{Z}^+$ on every X^n, $n \in \mathbb{N}$ by setting $k(x_1, x_2, \dots, x_n) = (T^k x_1, T^k x_2, \dots, T^k x_n)$ and also on every group $D^n(X, A)$ via $(kf)(x_0, x_1, \dots, x_n) = f(T^k x_0, T^k x_1, \dots, T^k x_n)$ for $f \in D^n(X, A)$. Thus every group $D^n(X, A)$ becomes a $\mathbb{Z}^+$-module and it is evident that the maps d^n are $\mathbb{Z}^+$-homomorphisms, so that for every q, $p \in \mathbb{Z}^+$ we obtain the group $H^q(\mathbb{Z}^+, D^p(X, A))$ which is the q-th cohomology group of the module $D^p(X, A)$. This means that this doubly indexed array of groups is a cohomological invariant of the action of T on the space X. Moreover the differentials $d^p : D^p(X, A) \to D^{p+1}(X, A)$ induce maps which we denote by $d_1^{p,q}$ of $H^q(\mathbb{Z}^+, D^p(X, A))$ into $H^q(\mathbb{Z}^+, D^{p+1}(X, A))$. Thus for fixed q these groups form a complex in the index p and hence we can form the cohomology groups of this complex, namely $Ker(d_1^{p,q})/Im(d_1^{p-1,q})$ which we shall denote by $H_T^{p,q}(X, A)$ displaying thus the data X and T and the parameters chosen p, q and A. It is a question for future investigation which of these invariants have a dynamical meaning and how to compute them.

4 Representation of Polynomials by Infinite Matrices and Semilinearization

If F is an algebraic field we consider the ring $F[x]$ of polynomials over F and we introduce in $F[x]$ an operation $*$ called substitution defined by $(p * q)(x) = p(q(x))$ for $p(x)$ and $q(x) \in F[x]$. The elements of $F^{\mathbb{N}}$, i.e., the sequences (f_n) with entries f_n in F we shall consider as infinite column vectors of the linear vector space $F^{\mathbb{N}}$. If $p(x) \in F[x]$ we write $([p(x)]^n)$ to mean the infinite column vector with the top entry 1, then $p(x), \cdots$ where $[p(x)]^n$ means the n-th algebraic power of $p(x)$. If $p(x) = a_0 + a_1 x + \cdots + a_k x^k$ we associate with it the infinite matrix $M(p)$ so that the entry $M_{i,j}$ indicates the coefficient of x^j in the polynomial $[p(x)]^i$, i, $j \in \mathbb{Z}^+$. This means that we can write the vector equation $([p(x)]^n) = M(p)(x^n)$ which may serve also as a definition of the matrix $M(p)$. We note that every such matrix $M(p)$ has only a finite number of nonzero entries in each row so that the product $M(p)M(q)$ is well defined for arbitrary two polynomials $p(x)$ and $q(x) \in F[x]$. We now show that the resulting matrix $M(p)M(q)$ equals the matrix $M(p * q)$ associated to $p * q(x)$. Indeed, if we substitute in the equation

$$([p(x)]^n) = M(p)(x^n) \tag{1}$$

for x the polynomial $q(x)$ we obtain, using associativity of matrix product:

$$([(p * q)(x)]^n) = M(p)([q(x)]^n) = M(p)M(q)(x^n) \ . \tag{2}$$

Comparing this with the defining equation (1) we obtain that $M(p * q) = M(p)M(q)$. We note that this result also proves that the operation $*$ is associative since the matrix multiplication is, and the map: $p(x) \longrightarrow M(p)$ is evidently injective. So we have just proved that the set $F[x]$ relative to the operation $*$ is a semigroup and the map $p(x) \longrightarrow M(p)$ is an injective homomorphism. It should be noted that the matrix $M(p)$ acts on the full spaced $F^{\mathbb{N}}$ so we can also consider the action of the semigroup $F[x]$ on $F^{\mathbb{N}}$. If $p(x) \in F(x)$ we shall consider the powers $p^{(r)}(x)$ relative to the operation $*$ not to be confused with the algebraic powers which are written as $[p(x)]^r$. From equation (2) and by induction we obtain for every $r \in \mathbb{N}$ the relation

$$([p^{(r)}(x)]^n) = [M(p)]^r(x^n) \quad . \tag{3}$$

From this vector equation we obtain (comparing the second entries from top)

$$p^{(r)}(x) = \sum_{i=0}^{\infty}[M(p)]^r_{1,i}x^i \quad . \tag{4}$$

Suppose now that $X \subseteq F$ is a subset of F which is p-invariant, i.e., $x \in X$ implies that $p(x) \in X$. Then we have an abstract system (X,T) on our hands (we do not consider any topology on X) where the action of T is defined by $Tx = p(x)$. If we define the map $i : X \to F^{\mathbb{N}}$ by

$$i(x) = (x^n) \in F^{\mathbb{N}} \quad , \tag{5}$$

then the equation (1) shows that the linear system $(F^{\mathbb{N}}, M(p))$ linearizes the system (X,T) since the embedding $i : X \to F^{\mathbb{N}}$ is equivariant, i.e.,

$$M(p)i(x) = i(Tx) \quad . \tag{6}$$

The equation (3) says that the orbit $\{T^k x : k \in \mathbb{N}\}$ of a point $x \in X$ can be computed from the powers $[M(p)]^k$ of the matrix $M(p)$. This result highlights the "trade-off" between linear and nonlinear mathematics but in an infinite dimensional space. We shall modify this trade-off by not giving up finite dimensionality and sacrificing instead the strict equality in the equation (6), replacing it by a less stringent condition and thereby introducing the notion of "semilinearization."

For the field F we shall now take the field of reals $\mathbb{R}$. Suppose $p(x) \in \mathbb{R}[x]$, $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and $X \subseteq \mathbb{R}$ is an interval invariant under $p(x)$. Thus setting again $Tx = p(x)$ we obtain a system (X,T). Suppose further that we can provide ourselves with real constants b_{ij} so that the following inequalities are verified on X:

$$\left.\begin{array}{rcl} p(x) & = & a_0 + a_1x + \cdots a_nx^n \\ [p(x)]^2 & \leq & b_{20} + b_{21}x + \cdots b_{2n}x^n \\ [p(x)]^n & \leq & b_{n0} + b_{n1}x + \cdots b_{nn}x^n \end{array}\right\} \quad . \tag{7}$$

Letting M denote the $(n+1)\times(n+1)$ matrix the rows of which are:

$(1,0,0,0,\dots,0)$, $(a_0, a_1, \dots, a_n)$, $(b_{20}, b_{21}, \dots, b_{2n}), \dots, (b_{n0}, b_{n1}, \dots, b_{nn})$

we may rewrite the system (7) as

$$([p(x)]^k) \leq M(x^k) \tag{8}$$

where $\leq$ now means the usual partial order relation between vectors in E^{n+1}. If we embed the interval X in E^{n+1} by $i(x) = (1, x, \dots, x^n) \in E^{n+1}$ the inequality (8) can be presented as

$$i(Tx) \leq Mi(x) \ . \tag{9}$$

Comparing (6) and (9) it is now clear what we mean by semilinearization. In (6) the matrix M is infinite and in (9) it is finite but instead of the equality we have only inequality. The question now is whether the relation (9) can give us some information concerning the orbits $\{T^i : i \in \mathbb{N}\}$ in X. Substituting for x in (9) its T-image Tx we obtain:

$$i(T^2x) \leq Mi(Tx) \ . \tag{10}$$

We shall now assume that the matrix M has nonnegative entries so that multiplication by M preserves the relation $\leq$ between vectors, i.e., $x \leq y$ implies $Mx \leq My$ for $x, y \in E^{n+1}$. If we now multiply (9) by M and compare with (10) we obtain that $i(T^2x) \leq M^2 i(x)$ and by induction we obtain

$$i(T^r x) \leq M^r i(x), r \in \mathbb{N} \ . \tag{11}$$

This result gives meaning to the above mentioned trade-off and we may say that the linear system (E^{n+1}, M) semilinearizes (X, T). Compared with equation (3) the relation (11) gives only a one-sided estimate of the orbit $\{T^r x : r \in \mathbb{N}\}$ in terms of powers M^r of the matrix M, but since M is finite we can use methods of linear algebra to compute M^r. We shall apply the powerful Cayley–Hamilton theorem to obtain M^r for $r \in \mathbb{N}$ from the characteristic polynomial $\varphi(\lambda)$ of the matrix M and its eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_{n+1}$.

It should be noted that the relation (11) was obtained assuming all the coefficients a_i, and b_{if} in the system (7) to be nonnegative. We now free ourselves, at least partially, from this restriction.

Suppose we contemplate a system of inequalities as in (7) where the top equation may be replaced by an inequality $p(x) \geq a_0 + a_1 x + \dots + a_n x^n$

with $p(x)$ being a continuous function and where some of them may have an opposite sense, i.e., $\geq$ instead $\leq$ and where some of the coefficients may be negative. It may happen that by multiplying some of these inequalities by -1 and by suitable choice of numbers $\epsilon_i = +$ or -1 one may achieve that the corresponding vector form of this system, analogous to (8) is $(\epsilon_k[p(x)]^k) \leq M(\epsilon_k x^k)$ where the matrix M has nonnegative entries. This means that the conclusion concerning the orbits $\{T^n x : n \in \mathbb{N}\}$ of the system (X,T) given by relation (11) remains valid if the topological embedding $i : X \to E^{n+1}$ is given by $i(x) = (1, \epsilon_1 x, \dots, \epsilon_n x^n)$. From this follows that for $T^r x = p^{(r)}(x)$ (the r-th iterate of $p(x)$) we obtain the estimate $\epsilon_1 p^{(r)}(x) \leq \sum_{i=0}^{n} \epsilon_i [M^r]_{1i} x^i$ for $r \in \mathbb{N}$ where $[M^r]_{ij}$ denotes the ij-th entry of the matrix M^r. Before giving some examples to illustrate the idea of semilinearization and to show its use in nonlinear analysis, we now formulate what we found so far.

Theorem 4.1. *Let (X,T) be a system on a finite dimensional space X and suppose there is a topological embedding $i : X \to E^n$ and an $n \times n$-matrix M with nonnegative entries such that $i(Tx) \leq Mi(x)$ for every $x \in X$. Then for every $r \in \mathbb{N}$ we have $i(T^r x) \leq M^r i(x)$. (The analogous statement is valid for the opposite relation $\geq$).*

Proof. By induction on r. For $r = 1$ it is assumed. Applying to both sides of $i(Tx) \leq Mi(x)$ the matrix M we obtain $Mi(Tx) \leq M^2 i(x)$ using the fact that M is nonnegative. Substituting in the first relation Tx for x we obtain $i(T^2 x) \leq Mi(Tx)$ which combined with the second relation yields $i(T^2 x) \leq M^2 i(x)$. Continuing in this manner we obtain the stated result. ■

The geometrical and practical meaning of the inequality $i(T^r x) \leq M^r i(x)$ is that it gives an estimate of the point $T^r x$ in X by the position of the corresponding vector $M^r i(x)$ in E^n. For computing the powers M^r of M we need an efficient method, and this is provided by the Cayley–Hamilton theorem. If $\varphi(\lambda)$ is the characteristic polynomial of the matrix M and all its roots $\lambda_1, \dots, \lambda_n$ are distinct, we consider the polynomial λ^r divided by $\varphi(\lambda)$ to obtain

$$\lambda^r = q(\lambda)\varphi(\lambda) + c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} \tag{12}$$

where $q(\lambda)$ is the quotient and $c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1}$ the remainder containing the constants $c_0, \dots, c_{n-1}$. For these constants we obtain n equations:

$$\lambda_i^r = c_0 + c_1\lambda_i + \cdots + c_{n-1}\lambda_i^{n-1} \tag{13}$$

substituting in (12) for λ the values λ_i for $i = 1, 2, \dots, n$. In the case of multiple roots the derivatives of the equation (12) will provide additional information so that for the unknown constant $c_0, \dots, c_{n-1}$ we obtain again n independent equations. We now illustrate this procedure on some examples.

Example 4.1. We investigate the iterations $\sin^{(n)} x$ of the function $\sin x$, i.e., $\sin^{(2)} x = \sin(\sin x), \dots, \sin^{(n+1)} x = \sin^{(n)}(\sin x)$. From the Taylor expansion $\sin x = x - a_3x^3 + a_5x^5 + \cdots$ where $a_n = (n!)^{-1}$ we obtain that

$$\sin x \leq x \quad \text{for} \quad x \geq 0 \ , \tag{14}$$

$$\sin x \geq x - a_3x^3 \quad \text{for} \quad x \geq 0 \tag{15}$$

(it is well known that these inequalities are valid on $[0,\infty)$.)

Multiplying by -1 and raising to power 3 we obtain $-[\sin x]^3 \geq -x^3$ so that combining with (15) we can write the vector relation

$$\begin{bmatrix} \sin x \\ -[\sin x]^3 \end{bmatrix} \geq \begin{bmatrix} 1, & a_3 \\ 0, & 1 \end{bmatrix} \begin{bmatrix} x \\ -x^3 \end{bmatrix} \ . \tag{16}$$

The characteristic polynomial of the matrix $M = \begin{bmatrix} 1, & a_3 \\ 0, & 1 \end{bmatrix}$ is $(1-\lambda)^2$ so that it has the root $\lambda = 1$ with multiplicity 2. The equation (12) is in this case $\lambda^r = q(\lambda)(1-\lambda)^2 + c_0 + c_1\lambda$ and its derivative $r\lambda^{r-1} = [q(\lambda)(1-\lambda)^2]' + c_\prime$, so that for the constants c_0 and c_1 we obtain two equations

$$\left.\begin{array}{rcr} 1 & = & c_0 + c_1 \\ r & = & c_1 \end{array}\right\} \ , \tag{17}$$

thus $c_1 = r$ and $c_0 = 1-r$. Substituting the matrix M for λ in the equation $\lambda^r = q(\lambda)(1-\lambda)^2 + c_0 + c_1\lambda$ we obtain, using the Cayley–Hamilton theorem that

$$M^r = (1-r)I + rM \quad \text{for} \quad r \in \mathbb{N} \tag{18}$$

where I is the identity matrix. From the general result (11) applied to our case $Tx = \sin x$ and (16) we obtain the vector relation

$$\begin{bmatrix} \sin^{(r)} x \\ -[\sin^{(r)} x]^3 \end{bmatrix} \geq M^r \begin{bmatrix} x \\ -x^3 \end{bmatrix} \ . \tag{19}$$

Computing from (18) the top row of the matrix M^r we obtain from (19) the inequality

$$\sin^{(r)} x \geq x - \frac{r}{6}x^3 \quad \text{for} \quad x \geq 0 \quad \text{and} \quad r \in \mathbb{N} \ . \tag{-19}$$

This inequality gives us an estimate of $\sin^{(r)} x$ from below. In order to obtain more accurate estimates of $\sin^{(r)} x$ and also estimates from above, we must take more members in the expansion of $\sin x$ as the initial inequality. In the next example we start with the inequality $\sin x \leq x - a_3x^3 + a_5x^5$. It should be noted that the inequalities

$$\left.\begin{array}{l} \sin x \leq \sum\limits_{k=0}^{n}(-1)^k a_{2k+1}x^{2k+1}, \quad n \text{ even} \\ \sin x \geq \sum\limits_{k=0}^{n}(-1)^k a_{2k+1}x^{2k+1}, \quad n \text{ odd} \end{array}\right\} \tag{20}$$

are valid on the whole interval $[0, \infty)$ which can be easily proved by induction on n.

Example 4.2. Choosing in (20) for $n = 2$ we obtain

$$\sin x \leq x - a_3x^3 + a_5x^5 \quad \text{on} \quad [0, \infty) \ . \tag{21}$$

Multiplying the inequality (14) by -1 and raising it to power 3 we obtain

$$\begin{aligned} -(\sin x)^3 \leq -(x - a_3x^3)^3 \quad &= -x^3 + 3a_3x^5 - 3a_3^2x^7 + a_3^3x^9 \\ &= -x^3 + 3a_3x^5 + a_3^2x^7(-3 + a_3x^2) \ , \end{aligned}$$

and since the expression $-3 + a_3x^2$ is ≤ 0 on the interval $[0, 3\sqrt{2}]$ we obtain the second inequality

$$-[\sin x]^3 \leq -x^3 + 3a_3x^5 \tag{22}$$

valid on the interval $[0, 3\sqrt{2}]$. Combining (21), (22) together with $[\sin x]^5 \leq x^5$ we obtain the vector relation

$$\begin{bmatrix} \sin x \\ -[\sin x]^3 \\ [\sin x]^5 \end{bmatrix} \leq \begin{bmatrix} 1, & a_3, & a_5 \\ 0, & 1, & 3a_3 \\ 0, & 0, & 1 \end{bmatrix} \begin{bmatrix} x \\ -x^3 \\ x^5 \end{bmatrix} \ . \tag{23}$$

As in Example 4.1. we must find the powers M^r of the matrix

$$M = \begin{bmatrix} 1, & a_3, & a_5 \\ 0, & 1, & 3a_3 \\ 0, & 0, & 1 \end{bmatrix} \ .$$

Its characteristic polynomial is $(1-\lambda)^3$ so its root $\lambda = 1$ has multiplicity 3. The equation (12) is in this case

$$\lambda^r = q(\lambda)(1-\lambda)^3 + c_0 + c_1\lambda + c_2\lambda^2 \ .$$

Its first and second derivative are:

$$\begin{aligned} r\lambda^{r-1} &= [q(\lambda)(1-\lambda)^3]' &+& \quad c_1 + 2c_2\lambda \quad \text{and} \\ r(r-1)\lambda^{r-2} &= [q(\lambda)(1-\lambda)^3]'' &+& \quad 2c_2 \ . \end{aligned}$$

The Cayley–Hamilton theorem yields

$$M^r = c_0 I + c_1 M + c_2 M^2 \tag{24}$$

where for c_0, c_1, c_2 we have the three equations

$$\left.\begin{aligned} 1 &= c_0 + c_1 + c_2 \\ r &= c_1 + 2c_2 \\ r(r-1) &= 2c_2 \end{aligned}\right\} \ . \tag{25}$$

From (24) and (25) it follows that the top row of the matrix M^r is: $(1, \frac{1}{6}r, \frac{1}{120}r(5r-4))$ so that the resulting estimate of $\sin^{(r)} x$, valid in the interval $[0, 3\sqrt{2}]$, is

$$\sin^{(r)} x \leq x - \frac{r}{6}x^3 + \frac{r}{120}(5r-4)x^5 \tag{26}$$

for $r \in \mathbb{N}$. Combining this result with that obtained in Example 4.1. we obtain a twosided estimate valid on the interval $[0, 3\sqrt{2}]$:

$$x - \frac{r}{6}x^3 \leq \sin^{(r)} x \leq x - \frac{r}{6}x^3 + \frac{r}{120}(5r-4)x^5 \tag{27}$$

for every $r \in \mathbb{N}$.

This result portrays how slowly the sequence $\sin^{(r)} x$ converges to zero as $r \to \infty$. If we set $x = \frac{1}{\sqrt{r}}$ we obtain from (27) a twosided estimate

$$\frac{5}{6}\frac{1}{\sqrt{r}} \leq \sin^{(r)} \frac{1}{\sqrt{r}} \leq \frac{1}{\sqrt{r}}(\frac{7}{8} - \frac{1}{30r}) \tag{28}$$

for every $r \in \mathbb{N}$.

In these examples the domain X of our system (X, T) was a one-dimensional interval. In the last example we show that the process of semi-linearization can be extended to higher dimensions.

Example 4.3. We consider the general quadratic transformation $T : E^2 \to E^2$ of the plane E^2 into itself i.e., T is defined by $T(x, y) = (x_1, y_1)$ for $(x, y) \in E^2$ where

$$\left.\begin{array}{l} x_1 = ax + by + \alpha x^2 + \gamma xy + \beta y^2 \\ y_1 = cx + dy + \lambda x^2 + \nu xy + \vartheta y^2 \end{array}\right\} . \tag{29}$$

The orbit $\{T^r(x,y) : r \in \mathbb{N}\}$ through the point (x,y) is obtained by substituting in (29) the point (x_1,y_1) for (x,y), then $(x_2,y_2) = T(x_1,y_1)$ for $(x_1,y_1)\cdots$ etc. We may imagine how complicated and algebraically involved a process it would be even for a very modest $r \in \mathbb{N}$. If we need some numerical information about the orbits near to origin $(0,0)$ we may adopt the policy of neglecting the higher powers than 2 in the process of iterations. This is precisely what we shall now consider. We restrict our attention to the first quadrant, i.e., the set $X \subseteq E^2$ is defined: $X = \{(x,y) : x \geq 0, y \geq 0\}$, and we assume that all ten coefficients $a, b, c, d, \alpha, \beta, \gamma, \lambda, \vartheta, \nu$ are ≥ 0. From (29) we obtain estimates of (x_1^2, x_1y_1, y_1^2) of the form:

$$\left.\begin{array}{rcl} x_1^2 & \geq & a^2x^2 + 2abxy + b^2y^2 \\ x_1y_1 & \geq & acx^2 + (ad+bc)xy + bdy^2 \\ y_1^2 & \geq & c^2x^2 + 2cdxy + d^2y^2 \end{array}\right\} . \tag{30}$$

Combining (29) and (30) we can write the following vector relation:

$$\begin{bmatrix} x_1 \\ y_1 \\ x_1^2 \\ x_1y_1 \\ y_1^2 \end{bmatrix} \geq \begin{bmatrix} a & b & \alpha & \gamma & \beta \\ c & d & \lambda & \nu & \vartheta \\ 0 & 0 & a^2 & 2ab & b^2 \\ 0 & 0 & ac & (ad+bc) & bd \\ 0 & 0 & c^2 & 2cd & d^2 \end{bmatrix} \begin{bmatrix} x \\ y \\ x^2 \\ xy \\ y^2 \end{bmatrix} . \tag{31}$$

This result suggests that our system (X,T) can be semilinearized in E^5 if we embed X into E^5 by the map

$$i(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} x \\ y \\ x^2 \\ xy \\ y^2 \end{bmatrix} \in E^5 . \tag{32}$$

Relation (31) says that the hypothesis of Theorem 4.1. is satisfied and since the matrix M of this system is nonnegative the conclusion of this theorem is valid in this case, yielding the onesided estimate for the points in orbits in the form

$$i\left(T^r \begin{bmatrix} x \\ y \end{bmatrix}\right) \geq M^r i \begin{bmatrix} x \\ y \end{bmatrix} . \tag{33}$$

For computing the powers M^r of M we would use the same procedure as explained in the previous examples.

References

[1] P.C. Baayen and J. deGroot. Linearazation of Locally Compact Transformation Groups in Hilbert Space, *Math. System Theory*, **Vol. 2, No. 4**, 363–379.

[2] A.H. Copeland and J. deGroot. Linearizations of Autohomeomorphisms, *Notices of Am. Math. Soc.*, **6**, (1959), 844.

[3] A.H. Copeland and J. deGroot. Linearization of a Homeomorphism, *Am. Math.*, **144** (1961), 80–92.

[4] M. Edelstein. On the Representation of Mappings of Compact Metrizable Spaces as Restrictions of Linear Transformations, *Canad. J. Math.*, **22** (1970), 372–375.

[5] M. Edelstein. On the Homomorphics and Isomorphic Embeddings of a Semiflow into a Radial Flow, *Pacific J. Math.*, **91** (1980), 281–291.

[6] J. deGroot. Every Continuous Mapping is Linear, *Notices Am. Math. Soc.*, **6** (1959), 754.

[7] L. Janos. A Converse of Banach's Contraction Theorem, *Proc. Amer. Math. Soc.*, **18** (1967), 287–289.

[8] L. Janos. On Representation of Self-Mappings, *Proc. Amer. Math. Soc.*, **26** (1970), 529–533.

[9] L. Janos. A Linearization of Semiflows in the Hilbert Space, ℓ_2, *Topology Preceedings*, Baton Rouge, LA, **2** (1977), 219–232.

[10] L. Janos, H.M. Ko, K.K. Tau. Edelstein's Contractivity and Attractors, *Proc. Amer. Math. Soc.*, **76** (1979), 339–344.

[11] L. Janos, R.C. McCann and J.L. Solomon. Nonexpansive Uniformly Asymptotically Stable Flows are Linear, *Canad. Math. Bull.*, **24(4)** (1981), 401–407.

[12] L. Janos. Differential Equations which are Topologically Linear, *Proc. Amer. Math. Soc.*, **83** (1981), 629–632.

[13] P.R. Meyers. A Converse to Banach's Contraction Theorem, *J. Res. Nat. Bureau of Standard*, **71B** (1971), 73–76.

[14] C. Moore and K. Schmidt. Coboundaries and Homomorphisms for Nonsingular Actions and a Problem of H. Helson, *Proc. London. Math. Soc.*, **40** (1980), 443–475.

[15] R. Nussbaum. Some Asymptotic Fixed Point Theorems, *Trans. Amer. Math. Soc.*, **171** (1972), 349–375.

[16] V.J. Opoitsev. A Converse to the Principle of Contracting Maps, *Russian Math. Surveys*, **31** (1976), 175–204.

[17] W. Parry and S. Tuncel. *Classification Problems in Ergodic Theory*, London Math. Soc. Lecture Note Ser. 67, Cambridge Univ. Press, Cambridge, 1982.

[18] J.A. Rus. Weakly Picard Mappings, *Comment. Math. Univ. Carolin.*, **34** (1993), 769-773.

Ludvik Janos
Department of Mathematics and Computer Science
Kent State University
Kent, OH 44242-0001 USA
janos@mcs.kent.edu

Finite Ultrametric Spaces and Computer Science

Vladimir A. Lemin

ABSTRACT The purpose of the paper is to describe a few properties of ultrametric spaces (in particular, of finite ones) and to demonstrate some applications of these properties to computer science.

A metric space (X, d) is called *ultrametric* [6] (or *non-Archimedean* [4], or *isosceles* [9]) if its metric satisfies the strong triangle axiom:

$$d(x, z) \leq max[d(x, y), d(y, z)]. \qquad (\Delta)$$

This is usually called the Ultrametric Axiom. Ultrametric spaces were described up to homeomorphism in [3, 21], up to uniform equivalence in [10], and up to isometry in [9, 20]. A survey of their metric [9, 20], geometric [14, 20], uniform [10], and categorical [11–17] properties can be found in the literature. The theory of ultrametric spaces is closely connected with various branches of mathematics. These are number theory (rings $\mathbf{Z}_p$ and fields $\mathbf{Q}_p$ of p-adic numbers), algebra (non-Archimedean normed fields), real analysis (the Baire space $B_{\aleph_o}$), general topology (generalized Baire spaces B_τ), p-adic analysis (field Ω), p-adic functional analysis (algebras of Ω-valued functions), lattice theory [17], Lebesgue measure theory [18], Euclidean geometry [14], category theory and topoi [13, 15, 16], and so on. These relations deal with infinite ultrametric spaces (mainly separable). For applications in computer science, finite spaces are of interest as well.

AMS Subject Classifications (2000): 54E35, 18B30, 54A25, 54C25.

1 Rationalization of Ultrametrics

Suppose we are going to study a metric space (X, d) (or any other mathematical object supplied with a real valued function) using a computer. The key problem that should be solved first of all is the following. Is it possible to approximate the metric $d(x, y)$ by a rational valued (binary rational valued) metric $r(x, y)$ close to the initial metric $d(x, y)$ in a certain sense? Theorem 1 gives an affirmative answer to this question for a wide class of ultrametric spaces. Theorem 3 does it for the others.

Theorem 1. [19] *Let (X,d) be an ultrametric space and $|V| = |\{d(x,y)|x,y \in X\}| \leq \aleph_0$. Then for any $\epsilon > 0$ and any $K > 1$ there exists an ultrametric $r(x,y)$ over X such that*

a) (X,d) and (X,r) are homeomorphic,

b) (X,d) and (X,r) are uniformly equivalent,

c) an identity map $i : (X,d) \to (X,r)$ is non-expanding,

d) an inverse map $i : (X,r) \to (X,d)$ is K-Lipschitz, i.e., $r(x,y) \leq d(x,y) \leq K \cdot r(x,y)$,

e) a difference between d and r is at most ϵ, $d(x,y) \geq r(x,y) \geq d(x,y) - \epsilon$,

f) all values of new ultrametric $r(x,y)$ are binary rational,

g) the identity map $i : (X,d) \to (X,r)$ induces a map $i^* : \{d(x,y)|x,y \in X\} \to \{r(x,y)|x,y \in X\}$ and the latter is one-to-one.

The requirement $|V| \leq \aleph_0$ is connected with the following property of ultrametric spaces, which does not hold for general metric spaces.

Theorem 2. [19, 20] *For any ultrametric space (X,d) the set of values of its metric $V = \{d(x,y)|x,y \in X\}$ has cardinality no greater than its weight, $|V| \leq w(X)$.*

This shows that the requirement $|V| \leq \aleph_0$ holds, in particular, for all separable ultrametric spaces. On the other hand, if the set V is uncountable then obviously no one-to-one rationalization can exist. However, omitting the requirement g) we can prove the following theorem.

Theorem 3. [19] *For any ultrametric space (X,d) there exists a uniformly equivalent binary rational ultrametric $r(x,y)$ satisfying the statements a)–f) of Theorem 1.*

So we see that for any ultrametric space (X,d) its metric $d(x,y)$ can be rationalized arbitrarily closely in several reasonable senses simultaneously (up to topological equivalence, uniform equivalence, an arbitrary small ϵ-variation, etc.). Note that the assertion e) means that the ultrametrics d and r are close to each other in the metric of uniform convergence on $X \otimes X$, while the assertion d) means the same for the metric of logarithm uniform convergence: $0 \leq \log[d/r] = |\log d - \log r| \leq \log K$.

Unfortunately, the situation with general metric spaces (not ultrametric ones) is worse and more complicated. Moreover, in general, their metrics cannot be rationalized even up to homeomorphism. Actually, if a metric $d(x,y)$ is rational valued then, for any irrational s, a closed ball $B(x,s) = \{y|d(x,y) \leq s\}$ is open. Hence (X,d) is small-inductive zero-dimensional, $\operatorname{ind} X = 0$. If the stronger equality holds, $\operatorname{Ind} X = \dim X = 0$ then, in view of a Morita–de Groot theorem [3, 21], (X,d) admits a topologically equivalent ultrametric $\Delta(x,y)$. Theorem 3 then enables us to rationalize it. So small inductive zero-dimensionality $\operatorname{ind} X = 0$ is necessary and (large inductive) zero-dimensionality $\operatorname{Ind} X = 0$ is sufficient for a topological rationalizability of a space. Moreover, the last requirement is necessary and sufficient for a topological rationalizability of general metric space by ultrametrics. This follows from zero-dimensionality Ind $X = 0$ of any ultrametric space and the Morita–de Groot theorem. A criterion for uniform rationalizability of general metric spaces by ultrametrics can be found in [19]. This is large proximate zero-dimensionality $\mathrm{In}\delta X = 0$ of a space (see [10] for the definition). The last property coincides with proximate zero-dimensionality $\delta \dim X = 0$ in the sense of Smirnov [10, 27]. Recall that $\delta \dim X = n$ means that Smirnov's compactification σX of a proximity space X is n-dimensional, $\dim(\sigma X) = n$ (see [26, 27]). As mentioned above small inductive zero-dimensionality ind $X = 0$ is necessary and (large inductive) zero-dimensionality Ind $X = \dim X = 0$ is sufficient for a topological rationalizability of a space by general metrics. Thus the following problem naturally arises.

Problem 1. There exist metric spaces with non-equal dimensions, ind $X = 0$ and Ind $X > 0$ (e.g., Roy's space [25] or Mrowka's space [22]). *Is it possible to introduce in such a space a rational-valued metric $r(x, y)$ up to topological equivalence at least?*

2 Distance Function and Lattice of Balls in Ultrametric Spaces

If an ultrametric space (X, d) is finite, then all the requirements mentioned above are obviously satisfied. Moreover, in this case the set $V = \{d(x, y) | x \neq y \in X\}$ of non-zero values of the metric d has additional computational properties.

Theorem 4. *For any finite ultrametric space (X, d) consisting of $n + 1$ points, the set V of values of its metric contains at most n elements.*

Proof. For $n = 1$ and $|X| = n + 1 = 2$, $|V| = 1$. Let $n = 2$ and $|X| = n + 1 = 3$. It is well known [6] that the strong triangle inequality implies that any three points form an isosceles triangle with base no greater than the sides. Thus $|V| \leq 2$. Suppose the theorem is true for any k-point space, where $k \leq n$, and let $|X| = n + 1$. By Lemma 3 [14], $n + 1$ points of an ultrametric space can be enumerated in such a way that

$$\begin{aligned} \min\{d(a_i, a_j) | i \neq j\} &= d(a_0, a_1) \leq d(a_0, a_2) \leq \cdots \leq d(a_0, a_n) \\ &= \max\{d(a_i, a_j)\}. \end{aligned} \tag{$*$}$$

Let us examine the chain ($*$) from right to left looking for the first place where there is a strict inequality. Two different cases should be considered here.

Case 1. The inequalities ($*$) have the form

$$\begin{aligned}\min\{d(a_i,a_j)|i \neq j\} &= d(a_0,a_1) \leq d(a_0,a_2) \\ &\leq \cdots \leq d(a_0,a_{n-1}) < d(a_0,a_n) = \max\{d(a_i,a_j)\}.\end{aligned}$$

Then it follows from the axiom (Δ) that the point a_n is at the same distance from any point a_k for $k < n$, $d(a_0,a_n) = d(a_k,a_n)$. By the inductive assumption, the set $V_{n-1} = \{d(a_k,a_j)|k,j < n\}$ has cardinality no greater than $n-1$. Thus the set $V = V_{n-1} \cup \{d(a_0,a_n)\}$ has cardinality $\leq n$.

Case 2. The inequalities $(*)$ have the form

$$\begin{aligned}\min\{d(a_i,a_j)|i \neq j\} &= d(a_0,a_1) \leq \cdots \leq d(a_0,a_{m-1}) < d(a_0,a_m) \\ &= \cdots = d(a_0,a_n) = \max\{d(a_i,a_j)\}.\end{aligned}$$

Then axiom (Δ) implies that all points a_k for $k = m, m+1, \ldots, n$, are at the same distance $d(a_0,a_n)$ from any point a_j for $j = 0,1,\ldots,k-1$. By the inductive assumption, the sets $V_{m-1} = \{d(a_k,a_j)|k,j < m\}$ and $V_m = \{d(a_k,a_j)|k,j \geq m\}$ have cardinalities $|V_{m-1}| \leq m-1$ and $|V_m| \leq (n-m+1)-1 = n-m$. Since $V \subseteq V_{m-1} \cup V_m \cup \{d(a_0,a_n)\}, |V| \leq (m-1)+(n-m)+1 = n$. ■

For general metric spaces a potency of the set V satisfies the inequality $|V| \leq n(n+1)/2$ and increases quadratically as $n \to \infty$. For ultrametric spaces it does linearly. This improves estimation of computer memory capacity and rate of computation.

A few other properties of ultrametric spaces follow from the structure of the set of balls of a space. This leads us to two results of different kinds connecting the theory of ultrametric spaces with lattice theory and computational modeling. To do it we, first of all, should refine the notion of radius of a ball. Usually a set $B(a,s) = \{x|d(x,a) \leq s\}$ is called a *closed ball of radius* s with a center located at a. However, such a notion gives us too many balls in a space (at least continuously many balls in each nonempty set). However, it seems more natural to say that there are only two balls in a one-point space $X = \{a\}$, namely, the empty set $\oslash$ and $\{a\}$, and only four balls in a two-point space $X = \{a,b\}$, namely, the empty set $\oslash$, two balls of radius zero (= points a and b), and the whole space X (= the ball of radius $d(a,b)$ with a center located at a or b). That is why we call a number s, written above, to be *a nominal radius* of a ball and introduce the following definition.

An actual radius of ball $B(a,s)$ is a number $r = \sup\{d(a,x)|x \in B(a,s)\}$.

Obviously, $r \leq s$. In an ultrametric space, balls have a lot of surprising properties, *exempli gratia,*

- *Any point of a ball is its center, i.e.,* $B(a,r) = B(x,r)$ *for any point* $x \in B(a,r)$.

- *An actual radius of a ball is equal to its diameter, i.e.,* $r = \sup\{d(x,y)|x,y \in B(a,s)\}$.
- *Any two balls are either disjoint or one of them is a subset of the other.*
- *If the balls* $B(a,s)$ *and* $B(b,t)$ *are disjoint, then* $d(a,b) = d(x,y)$ *for any* $x \in B(a,s)$ *and any* $y \in B(b,t)$.

Moreover, it turns out that the set of balls of an ultrametric space is a lattice $L(X)$ and there is a duality between ultrametric spaces and a certain class of lattices.

Theorem [17]. *For any ultrametric space* (X,d) *there is a complete, atomic, tree-like, and real graduated lattice* $(\mathbf{L}(X),\ \sup, \cap, r(B(\alpha)))$ *and for any complete, atomic, tree-like, and real graduated lattice* $(L, \vee, \wedge, r(\alpha))$ *there is an ultrametric space* $(\mathbf{A}(L),\ \Delta)$ *such that*

- *the space* (X,d) *is isometric to the space* $(\mathbf{A}(\mathbf{L}(X)),\ \Delta)$*;*
- *the lattice* $(L, \wedge, \vee, r(\alpha))$ *is isomorphic to the lattice* $(\mathbf{L}(\mathbf{A}(L)),\ \cap, \sup, r(B(\alpha)))$.

A similar theorem holds for morphisms. This means that there is an isomorphism functor between the category **ULTRAMETR** of ultrametric spaces and non-expanding maps and the category **LAT**$*$ of complete, atomic, tree-like, and real graduated lattices and isotonic, semi-continuous, non-extensive maps (see [17] for proofs, definitions and details).

Theorem 5. *For any finite ultrametric space* (X,d) *consisting of* n *points, the set* $\mathbf{L}(X)$ *of its closed balls contains at most* $2n$ *elements,* $|\mathbf{L}(X)| \leq 2|X|$.

Proof. For $n = 1, X = \{a\}$ is a singleton and $\mathbf{L}(X) = \{\emptyset, \{a\}\}$, thus $|\mathbf{L}(X)| = 2 = 2|X|$. Suppose the theorem holds for any k-point space with $k \leq n$ and let $|X| = n+1$. Following similar arguments as above we enumerate the points of space in the same manner and consider two cases.

Case 1. $\min\{d(a_i,a_j)|i \neq j\} = d(a_0,a_1) \leq d(a_0,a_2) \leq \cdots \leq d(a_0,a_{n-1}) < d(a_0,a_n) = \max\{d(a_i,a_j)\}$. Denote by X' the set $X' = \{a_0,a_1,a_2,\ldots,a_{n-1}\}$. By the inductive assumption, $|\mathbf{L}(X')| \leq 2n$. Since $d(a_0,a_n) = d(a_k,a_n) > d(a_k,a_m) \forall k,m < n$, we have $\mathbf{L}(X) = \mathbf{L}(X') \cup X$, thus $|\mathbf{L}(X)| \leq 2n+1 < 2(n+1) = 2|X|$.

Case 2. $\min\{d(a_i,a_j)|i \neq j\} = d(a_0,a_1) \leq \cdots \leq d(a_0,a_{m-1}) < d(a_0,a_m) = \cdots = d(a_0,a_{n-1}) = d(a_0,a_n) = \max\{d(a_i,a_j)\}$. Denote by X_{m-1} and X_m the sets $X_{m-1} = \{a_0,a_1,a_2,\ldots,a_{m-1}\}$ and $X_m = \{a_m,a_{m+1},\ldots,a_n\}$. By the inductive assumption, $|\mathbf{L}(X_{m-1})| \leq 2|X_{m-1}| = 2m$ and $|L(Xm)| \leq 2|X_m| = 2(n-m+1)$. As mentioned above all points a_k for $k \geq m$, are at the same distance $d(a_0, a_n)$ from all points a_j for $j < m$. Thus $\mathbf{L}(X) \subseteq \mathbf{L}(X_{m-1}) \cup \mathbf{L}(X_m) \cup \{X\}$. The empty set $\oslash$ belongs to both of the lattices $\mathbf{L}(X_{m-1})$ and $\mathbf{L}(X_m)$ and is counted in both of them. Therefore $|\mathbf{L}(X)| \leq 2m + 2(n-m+1) - 1 + 1 = 2(n+1) = 2|X|$. ■

Example. Let $X = \{0, 1, 2, \ldots, n-1\}$ be a subset of natural numbers equipped with the following ultrametric $d(k, m) = \max(k, m)$. The set $\mathbf{L}(X)$ consists of n singletons $\{k\}$ (= balls of radius zero), the empty set $\oslash$, and $n-1$ balls $B_m = \{0, 1, 2, \ldots, m\} = B(0, m)$ of radius m for $m = 1, 2, \ldots, n-1$. Thus $|\mathbf{L}(X)| = 2n$ and the estimation $|\mathbf{L}(X)| \leq 2|X|$ cannot be improved.

Note. The inequality $|\mathbf{L}(X)| = 2|X|$ does not hold for infinite spaces. Let $X = \mathbf{Q}_+$ be a set of non-negative rationals with the same metric $d(x, y) = \max(x, y)$ (we call it *the max-metric*). The set $\mathbf{L}(X) = \{B(0, r) | r \in \mathbf{R}_+\}$ has cardinality of the continuum.

For general metric spaces a potency $|\mathbf{L}(X)|$ of the set of balls increases quadratically as $n \to \infty$. For ultrametric spaces it increases linearly as the cardinality $|V|$ does. Since all closed balls $B(a, s) = \{x | d(x, a) \leq s\}$ are open they form a clopen base for the topology of X. This enables us to construct a computational model for a space X in the sense of Lawson [8]. Following this way Flagg and Kopperman studied a problem of existence of an **algebraic** computational model and proved the following.

Theorem [2] *A topological space X has an algebraic computational model if and only if it is separable, complete and ultrametric* (see [8, 2] and references there).

3 Categorical Operations and Embedding Theorems

Let us consider a category **METR** ($\mathbf{METR_c}$) of all metric spaces (of diameter not greater than c) and non-expanding maps. These are the maps $f : (X, d) \to (Y, r)$, which do not enlarge distances, i.e., $r(f(x),\ f(y)) \leq d(x, y)$ for all x and y in X. It is known that sums and products, pull-backs and push-outs, equalizers and co-equalizers, and limits of direct and inverse spectra exist in $\mathbf{METR_c}$ and that the subcategory of ultrametric spaces and the same maps $\mathbf{ULTRAMETR_c}$ is closed in $\mathbf{METR_c}$ with respect to these operations (see [11, 15, 17]). As for the category **METR** we have the following.

Proposition 1. *A sum of objects does not exist in* **METR** *even for two singletons.*

Proposition 2. *For any finite family* $\{(X_k, d_k)|k = 1, 2, \ldots, n\}$ *of metric spaces, there exists a product* (X_Π, d_Π) *of these spaces in* **METR** *(called* a metric product*)* $m\Pi\{(X_k, d_k)\} = (X_\Pi, d_\Pi)$, *where* $X_\Pi = \Pi X_k$ *is a product of the sets* X_k *in the category* **SET**, *and* $d_\Pi(\{x_k\}, \{y_k\}) = \max\{d_k(x_k, y_k)|k = 1, 2, \ldots, n\}$.

Proposition 3. *A metric product of an infinite family* $\{(X_\alpha, d_\alpha)|\alpha \in I\}$ *of metric spaces exists iff there is* $c > 0$ *such that for almost all of* (X_α, d_α), *diam* $(X_\alpha, di_\alpha) < c$.

Proof. The part "if" of the statement is obvious, $\{d_\Pi(\{x_\alpha\}, \{y_\alpha\}) = \sup d_\alpha(x_\alpha, y_\alpha)|\alpha \in I\}$ is a finite number for any pair of points $\{x_\alpha\}, \{y_\alpha\}$ in $m\Pi\{(X_\alpha, d_\alpha)\}$. Consider a family $\{(X_\alpha, d_\alpha)|\alpha \in I\}$, which does not satisfy the assertion of the proposition. Then there exists a subfamily $\{(X_n, d_n)|n \in \mathbb{N}\}$ and points $x_n, y_n \in X_n$, such that $d_n(x_n, y_n) \geq n$. Suppose that the product $m\Pi\{(X_\alpha, d_\alpha)\}$ does exist. Then there are non-expanding projections $p_\alpha : m\Pi\{(X_\alpha, d_\alpha)\} \to (X_\alpha, d_\alpha)$ satisfying the standard requirements. Let $Y = \{a\}$ be a singleton and let $f_\alpha, g_\alpha : Y \to (X_\alpha, d_\alpha)$ be two families of maps such that $f_n(a) = x_n, g_n(a) = y_n \forall n \in \mathbf{N}, f_a(a) = g_\alpha(a) = x_\alpha$ for $\alpha \in \mathbf{I} \backslash \mathbf{N}$. Since $m\Pi\{(X_\alpha, d_\alpha)\} = (X_\Pi, d_\Pi)$ is a product in the category **METR** there are the maps $f : Y \to X_\Pi$ and $g : Y \to X_\Pi$ such that $p_\alpha \cdot f = f_\alpha$ and $p_\alpha \cdot g = g_\alpha \forall \alpha \in \mathbf{I}$. Since projections p_α are non-expanding the distance $d_\Pi(f(a), g(a))$ should satisfy the inequality $d_\Pi(f(a), g(a)) \geq d_\alpha(p_\alpha f(a), p_\alpha g(a)) = d_\alpha(f_\alpha(a), g_\alpha(a)) \forall \alpha \in \mathbf{I}$, in particular, $d_\Pi(f(a), g(a)) \geq d_n(f_n(a), g_n(a)) = d_n(x_n, y_n) \geq n, \forall n \in \mathbf{N}$. This implies that $d_\Pi(f(a), g(a)) = \infty$. ∎

Note. If $\sup\{\text{diam}(X_\alpha, d_\alpha)\} = c < \infty$ then the product $m\Pi\{(X_\alpha, d_\alpha)\}$ in **METR** coincides with the product of the same spaces in $\mathbf{METR_c}$.

Lemma 1. *Subcategories* **ULTRAMETR** *and* $\mathbf{ULTRAMETR_c}$ *are closed in* **METR** *and* $\mathbf{METR_c}$ *with respect to product.*

Proof. The proof for $\mathbf{METR_c}$ can be found in [15] and that for **METR** is similar. ∎

Denote by (D_α, d_α) a space consisting of two real numbers $\{0, c_\alpha\}$ with a metric $d_\alpha(0, c_\alpha) = c_\alpha$. Since two-point spaces are obviously ultrametric and any subset of an ultrametric space is ultrametric again, we obtain the following.

Corollary 1. *Any subset of a metric product* $m\Pi\{(D_\alpha, d_\alpha)|\alpha \in I\}$ *of an arbitrary family of two-point spaces, is ultrametric.*

Theorem 6. *For any ultrametric space* (X, d) *there exists a family* $\{(D_\alpha, d_\alpha)|\alpha \in I\}$ *of two-point spaces such that* (X, d) *is isometric to a subspace of their metric product and a number of factors in the product is not greater than the weight of* $X, |I| \leq w(X)$.

Proof. For the sake of simplicity, suppose first that X is of finite diameter. Let $Z \subset X$ be a dense subset of cardinality τ, $\tau = |Z| = w(X)$. For any $z \in Z$, consider a set of all open balls $B^\circ(z,r) = \{x|d(z,x) < r\}$ of radius r such that there is $y \in X$ with $d(z,y) = r$ (= the balls with non-empty spheres). By Theorem 2, the set of all such balls is of cardinality at most τ. Denote by $\mathbf{L}^\circ(X)$ a set of all these balls for all $z \in Z$, i.e., $\mathbf{L}^\circ(X) = \{B^\circ(z,r)|z \in Z\} = \{B(\alpha)|\alpha \in I\}$. Clearly $|\mathbf{L}^\circ(X)| = |\mathbf{I}| = \tau \cdot \tau = \tau$ (here we consider spaces of infinite weight, for finite spaces see Theorem 9 below). For any such a ball $B^\circ(z,r) = B(\alpha)$, let us define a function $f_\alpha : X \to D_\alpha = \{0, c_\alpha\}$ as follows: $f_\alpha|_{B(\alpha)} = 0, f_\alpha|_{X \setminus B(\alpha)} = c_\alpha$, where $c_\alpha = r$ is a radius of the ball $B(\alpha) = B^\circ(z,r)$. Properties of balls mentioned in Section 2 imply that f_α's are non-expanding. A categorical product $f = \Pi f_\alpha$ is non-expanding as well and it maps X in the metric product $m\Pi\{D_\alpha|\alpha \in I\}$. For any two points x and y in X there is a point $z \in Z$ such that $x \in B^\circ(z,r)$, $y \notin B^\circ(z,r)$ and $r = d(x,y) = d(z,y)$. For any $\alpha, d_\alpha(f_\alpha(x), f_\alpha(y)) = c_\alpha$ if $x \in B(\alpha)$ and $y \notin B(\alpha)$ or $x \notin B(\alpha)$ and $y \in B(\alpha)$, and $d_\alpha(f_\alpha(x),\ f_\alpha(y)) = 0$ otherwise. Consequently, $d_\Pi(f(x), f(y)) = d_\Pi(\{f_\alpha(x)\},\ \{f_\alpha(y)\}) = \sup\{d_\alpha(f_\alpha(x),\ f\alpha(y))|\alpha \in I\} = \sup\{c_\alpha|c_\alpha \le d(x,y)\} = d(x,y)$. Thus f is an isometric embedding. ■

To embed a space of infinite diameter we are to study a category **ULTRAMETR*** (**METR***) of all (ultra-) metric spaces $(X, d(x,y), x^*)$ with a base point x^* (pointed metric spaces) and non-expanding maps that take a base point to a base point. In contrast to Proposition 1 we have

Proposition 4. *For any family* $\{(X_\alpha,\ d_\alpha,\ x_\alpha*)|\alpha \in I\}$ *of pointed metric spaces, there exists a sum* $(X_\sum,\ d_\sum,\ x^*)$ *of these spaces in the category* **METR*** $m\sum\{(X_\alpha,\ d_\alpha,\ x_\alpha*)|\alpha \in I\} = (X_\sum,\ d_\sum,\ x^*)$, *where* $(X_\sum,\ x^*)$ *is a sum of the sets* $(X_\alpha, x_\alpha*)$ *in the category* **SET***, *and* $d_\sum(x_\alpha,\ x^*) = d_\sum(x_\alpha, x_\alpha^*), d_\sum(x_\alpha, y_\alpha) = d_\alpha(x_\alpha, y_\alpha), d_\sum(x_\alpha, y_\beta) = d_\alpha(x_\alpha, x_\alpha^*) + d_\beta(y_\beta, x_\beta^*)$ *for* $\alpha \ne \beta$.

Proposition 5. *For any family* $\{(X_\alpha, d_\alpha, x_\alpha^*)|\alpha \in I\}$ *of pointed ultrametric spaces, there exists a sum* $(X_\sum, d_\sum, x^*)$ *of these spaces in the category* **ULTRAMETR*** *(called* a pointed sum*)* $m\sum\{(X_\alpha, d_\alpha, x_\alpha^*)|\alpha \in I\} = (X_\sum, d_\sum, x^*)$, *where* $(X_\sum, x^*)$ *is a sum of the sets* (X_α, x_α^*) *in the category* **SET***, *and* $d_\sum(x_\alpha, x^*) = d_\sum(x_\alpha, x_\alpha^*), d_\sum(x_\alpha, y_\alpha) = d_\alpha(x_\alpha, y_\alpha), d_\sum(x_\alpha, y_\alpha) = \max[d_\alpha(x_\alpha, x_\alpha^*), d_\beta(y_\beta, x_\beta^*)]$ *for* $\alpha \ne \beta$.

Proposition 6. *For any family* $\{(X_\alpha, d_\alpha, x_\alpha^*)|\alpha \in I\}$ *of pointed metric spaces, there exists a product* (X_Π, d_Π, x^*) *of these spaces in the category* **METR*** *(called a* pointed product*)* $m\Pi^*\{(X_\alpha, d_\alpha, x_\alpha^*)|\alpha \in I\} = (X_\Pi, d_\Pi, x^*)$. *Here* $x^* = \{x_\alpha^*\}, d_\Pi(\{x_\alpha\}, \{y_\alpha\}) = \sup\{d_\alpha(x_\alpha, y_\alpha)|\alpha \in I\}$, *and* X_Π *is a subset of the Cartesian product* ΠX_α *of the sets* X_α *consisting of those points* $\{x_\alpha\} \in \Pi X_\alpha$ *that are at a finite distance from the point* $x^*, \sup\{d_\alpha(x_\alpha, x_\alpha^*)|\alpha \in I\} < \infty$.

Lemma 2. *The subcategory* **ULTRAMETR*** *is closed in* **METR*** *with respect to multiplication* (see [15]).

Coming back to the proof of Theorem 6 we choose an arbitrary point $x^* \in X$ as a base point in X, choose 0 as a base point in every D_α, and modify slightly the definition of the functions $f_\alpha(x)$. Namely let $f_\alpha(x)$ be the same as above if $x^* \in B(\alpha)$ and let it be $c_\alpha - f_\alpha(x)$ otherwise. Then new functions $f_\alpha(x)$ take x^* to $0_\alpha \in D_\alpha$ for all α, the product $f = \Pi f_\alpha$ takes x^* to $\mathbf{0} = \{0_\alpha\} \in m\Pi^*\{D_\alpha\}$, and it maps (X, d, x^*) in $m\Pi^*\{(D_\alpha, d_\alpha, 0_\alpha)\}$ in view of the definition of the pointed product. The rest of the proof is obvious. ∎

Compare the last theorem with the classic theorems on embedding of topological spaces in a product of two-point spaces.

Theorem (P.Alexandroff, 1936 [1]). *Any T_0-space of weight $\leq \tau$ can be embedded homeomorphically in the Alexandroff cube F^τ of weight τ.*

Recall that the Alexandroff cube $F^\tau = \Pi\{F_\alpha | \alpha \in \mathbf{I}\}$ is a topological product (a product in the category **TOP**) of τ connected two-point spaces $F_\alpha = \{0, 1\}_\alpha$, i.e., the spaces with open sets $\oslash$, $\{0\}$, and $\{0, 1\}$.

Theorem (P.Alexandroff, 1936 [1]). *Any small inductive zero-dimensional space of weight $\leq \tau$ (*ind $X = 0$*) can be embedded homeomorphically in the Cantor cube D^τ of weight τ (= a topological product of τ discrete two-point spaces $D_\alpha = \{0, 1\}_\alpha$) .*

A close analogy between these theorems is obvious. The way that they differ is probably more interesting. First, the weight of spaces F^τ and D^τ is exactly τ whereas the weight of a metric product $m\Pi\{(D_\alpha, d_\alpha) | \alpha \in \mathbf{I}\}$ depends on the "sizes" of factors; it is generally greater, $\tau \leq w(m\Pi\{D_\alpha | \alpha \in I\}) \leq 2^\tau$. E.g., a product $m\Pi\{\{0, 1/n\} | n \in \mathbf{N}\}$ of a countable set of two-point spaces $\{0, 1/n\}$ is separable. It is easy to show a countable dense subset there. And the product $X_\Pi = m\Pi\{\{0, 1\}_n | n \in \mathbf{N}\}$ of a countable set of two-point spaces $\{0, 1\}$ is a metrically discrete set of cardinality of the continuum, $d(x, y) = 1 \forall x, y \in X_\Pi$. Thus its weight is also c. Furthermore, in Alexandroff's theorems, all factors are topologically equivalent to each other, their product is a universal space in the category **TOP**, whereas the factors $D_\alpha = \{0_\alpha, c_\alpha\}$ in Theorem 6 are not metrically equivalent to each other. Theorem 6 assigns, for each space (X, d), its own particular family of two-point spaces, and their product is not metrically universal.

The problem of the existence of a universal ultrametric space for all spaces of a given weight was studied in [9, 20]. For any cardinal τ, there is constructed an ultrametric space LW_τ, which contains isometrically all ultrametric space of weight $\leq \tau$ (a *τ-universal space*). However, sometimes this space turns out to be too large. In the rest of the section, we describe a smaller universal space $LV_\tau \subset LW_\tau$, relate the spaces LV_τ and LW_τ to the notions of metric sum and product, and deduce a theorem on embedding of ultrametric spaces in Banach spaces.

Definition [20]. Let τ be an arbitrary cardinal, $\omega(\tau)$ be the first ordinal of potency $\tau, W_\tau = \{\alpha|\alpha < \omega(\tau)\}$ be the set of all ordinals smaller than $\omega(\tau)$ (= the set of all ordinals of potency $< \tau$). A function $f : \mathbf{Q}_+ \to W_\tau$ is called *eventually* 0*-valued* (or *eventually vanished*) if there is a number $X(f)$ such that $f(x) = 0 \ \forall \ x > X$. For any two such maps f and g let us define a distance $\Delta(f,g)$ by the equality $\Delta(f,g) = \sup\{x|f(x) \neq g(x)\}$. Denote by (LW_τ, Δ) the space of all these functions equipped with the metric Δ.

It is easy to prove that Δ is a metric satisfying the strong triangle axiom and that (LW_τ, Δ) is a complete ultrametric space [9, 20].

Theorem 7. *Every ultrametric space of weight $\leq \tau$ can be embedded isometrically in LW_τ. The weight of LW_τ equals its potency and equals $\tau^{\aleph_0}, |LW_\tau| = w(LW_\tau) = \tau^{\aleph_0}$.*

Proof. The key idea of the proof is used in the proof of Theorem 8 below for $\tau = \aleph_0$. For arbitrary τ as well as for details see [20]. ■

For any cardinal $\tau \leq c, \tau^{\aleph_0} = c$. The weight of the universal space $w(LW_\tau) = \tau^{\aleph_0} = c$ seems to be too great, especially for finite τ. However, the following proposition shows that a weight of a τ-universal space cannot be smaller even for $\tau = 2$.

Proposition 7. *If an ultrametric space (U,d) contains isometrically all two-point spaces, then its weight is not less than that of the continuum, $w(U,d) \geq c$.*

Proof. It follows directly from Theorem 2. ■

So for all cardinals $\tau \leq c$, the weight of the universal space LW_τ is the smallest of possible ones. Cardinals $\tau > c$ can be divided in two classes: namely, satisfying $\tau^{\aleph_0} = \tau$ and $\tau^{\aleph_0} > \tau$, respectively. For all cardinals in the first class the weight of LW_τ is $\tau^{\aleph_0} = \tau$, thus it is the smallest weight of a τ-universal space. For cardinals from the second class we have the following problem [20].

Problem 2 [20]. Does there exist a τ-universal ultrametric space (U,d) with weight smaller than $\tau^{\aleph_0}$ for cardinals τ such that $\tau^{\aleph_0} > \tau > c$? In particular, does there exist one of weight τ?

A partial answer to this problem was given by J. Vaughan [31]. Analyzing the proof of Theorem 7 [20], he has mentioned that actually every ultrametric spaces of weight $\leq \tau$ is embedded there in a somewhat smaller space (denoted by $LW_\tau^\backprime$), that is a closure of the set of all "bounded" eventually 0-valued functions [31]. A function f is called *bounded* if it maps $\mathbf{Q}_+$ into a set $[0, \beta]$ for some $\beta < \omega(\tau)$. Vaughan proved that the weight of $LW_\tau^\backprime$ is $\tau \cdot \sum\{\alpha^{\aleph_0}|\alpha < \tau\}$, [31]. Thus for cardinals from the second class that satisfy the inequality $\sum\{\alpha^{\aleph_0}|\alpha < \tau\} \leq \tau$ (we call them *Vaughan's cardinals*), there exists a τ-universal space of weight τ, namely $LW_\tau^\backprime$ [31]. Vaughan also shows that, under the Singular Cardinal Hypothesis (a set-theoretic assumption, whose negation implies the existence of measurable cardinals), every cardinal $\tau > c$ is a Vaughan cardinal [31]. So the problem is solved in the affirmative under ZFC + SCH. However, the question is still open in ZFC without any additional set-theoretic axioms (see [20, 31 and 32] for details and discussion on the relation of Problem 2 to the problem of large cardinals).

In fact, as mentioned in [20], even the subset $LV_\tau \subset LW_\tau$ of all monotone left semi-continuous functions $f : \mathbf{Q}_+ \to W_\tau$ is enough to contain all spaces of weight $\leq \tau$. It is easily proved that $LV_\tau \subset LW^{\backslash}_\tau$. The difference between LW_τ and $LW^{\backslash}_\tau$ is inessential for finite τ (see Theorem 8 below), clearly $LW^{\backslash}_\tau = LW_{\tau-1}$, whereas that between LW_τ and LV_τ is great whenever $\tau \geq 2$. For cardinals of uncountable confinality $LW^{\backslash}_\tau = LW_\tau$ [31].

Consider the simplest case, $\tau = 2$. The space LV_2 consists of all monotone decreasing left semi-continuous functions $f : \mathbf{Q}_+ \to \{0, 1\}$. Every such a function is of the form $f(q) = f_t(q) = 1$ for $q \leq t, f_t(q) = 0$ for $q > t$. Thus it can be identified with the real number $t \in \mathbf{R}_+$ and the space LV_2 is none other than a set of positive numbers equipped with the max-metric, $(LV_2, \Delta) = \mathbf{R}_+, \max)$. The latter is an ultrametric space of weight c and it is obviously universal for all two-point spaces; any two-point space $\{0, d\}$ can be embedded in $(\mathbf{R}_+, \max)$ by the identity $i(0) = 0,\ i(d) = d$. Moreover, this is just the same as the pointed sum of all two-point spaces of positive "lengths".

Proposition 8. *The space* $(LV_2, \Delta) = (\mathbf{R}_+, \max)$ *is naturally isometric to the pointed sum of all two-point spaces of positive "lengths"* $(LV_2, \Delta) = m\sum^*\{\{0, r\} | r \in \mathbf{R}_+\}$.

To get a universal space of a type of metric product it is enough to take a pointed product of the same two-point spaces $m\Pi^*\{\{0, r\} | r \in \mathbf{R}_+\}$. There are continuously many factors in that product. Better, it is enough to multiply a countable number of factors to reach universality, namely, a countable set of two-point spaces whose lengths are dense in $\mathbf{R}_+$, e.g., spaces of rational lengths, $m\Pi^*\{\{0, q\} | q \in \mathbf{Q}_+\}$. We call this space a *rational cube*, and denote it by q^Q.

Proposition 9. *The rational cube* q^Q *is a* 2*-universal ultrametric space.*

Proof. An isometric embedding $i : \{0, d\} \to m\Pi^*\{\{0, q\} | q \in \mathbf{Q}_+\}$ defined as $i(0) = \mathbf{0} = \{0_q\}, i(d) = \{x_q\}$, where $x_q = q$ for $q \leq d, x_q = 0$ for $q > d$, is a desired one. ■

The space LW_2 is also 2-universal, by Theorem 7 (and also by the inclusion $LW_2 \supset LV_2$). The amazing fact is that these two spaces coincide with one another.

Proposition 10. *The universal ultrametric space* LW_2 *is naturally isometric to the rational cube* $q^Q = m\Pi^*\{\{0, q\} | q \in \mathbf{Q}_+\}$.

Proof. The desired isometry takes any function $f \in LW_2$ to a point $\{x_q\} \in \Pi\{\{0,q\}|q \in \mathbf{Q}_+\}$ defined by the equalities $x_q = q$ if $f(q) = 1, x_q = 0$ if $f(q) = 0$. Since f is eventually 0-valued there is $p > 0$ such that $f(q) = 0$ for all $q > p$. Thus $d_\Pi(\mathbf{0}, \{x_q\}) = \sup\{d_q(0, x_q)|q \in \mathbf{Q}_+\} = \sup\{q|x_q \neq 0\} = \sup\{q|f(q) = 1\} \leq \sup\{q|q \leq p\} = p$. Therefore the point $\{x_q\}$ belongs to the *pointed* product $m\Pi^*\{\{0,q\}|q \in \mathbf{Q}_+\} = q^Q$. On the contrary, an inverse map i^{-1} takes any point $\{x_q\} \in q^Q$ to a function $f : \mathbf{Q}_+ \to \{0,1\}$ defined as follows: $f(q) = 1$ if $x_q = q, f(q) = 0$ if $x_q = 0$. Since $\{x_q\}$ belongs to the *pointed* product, $d_\Pi(\mathbf{0}, \{x_q\}) = \sup\{d_q(0, x_q)|q \in \mathbf{Q}_+\} = \sup\{x_q|x_q \neq 0\} = \sup\{q|x_q \neq 0\}$ is finite. Thus there is $p > 0$ such that $x_q = 0 \ \forall q > p$, hence the corresponding function $f(q)$ eventually vanishes, $f(q) \in LW_2$. Isometry is obvious. ■

As mentioned above the weight of LW_τ as well as that of LV_τ is c for all τ from 2 to c. LV_2 is 2-universal but it is not n-universal for any $n > 2$, for $n = 3$ either; this follows, for instance, from the fact that there is no equilateral triangle in $LV_2 = (\mathbf{R}_+, \max)$. Similarly, LV_n is n-universal but not $(n+1)$-universal. Fortunately, the space LW_2 turns out to be τ-universal not only for every finite $\tau = n$, but also for all cardinals no greater than c. We prove this theorem below for separable spaces.

Theorem 8. *The space* $(LW_2, \Delta) = q^Q$ *is metrically universal for all separable ultrametric spaces.*

Proof. To prove the theorem, we follow [20] with a minor modification. Recall the method of embedding of a given separable ultrametric space (X, d) into $LW_{\aleph_o}$ (in fact, in $LV_{\aleph_o}$), [20]. First, we choose a countable dense subset $Y = \{a_0, a_1, \ldots, a_n, \ldots\}$ in X and define an isometric embedding $i : Y \to LV_{\aleph_o}$ inductively. Denote by $f_n(q)$ the image $i(a_n)$ and put $i(a_0) = f_0(q) \equiv 0$, and $f_1(q) = f_0(q)$ for $q > d(a_0, a_1)$ and $f_1(q) = 1$ for $q \leq d(a_0, a_1)$. Obviously, $\Delta(f_0, f_1) = d(a_0, a_1)$. This provides us with the base of induction. Suppose the points $a_0, a_1, \ldots, a_{n-1}$ are already embedded isometrically in $LV_{\aleph_o}$ in such a way that $f_k(\mathbf{Q}_+) \subseteq \{0, 1, \ldots, k\}$ for $k < n$. To embed the point a_n, we compute $\min\{d(a_n, a_k)|k < n\} = d_n$, take a point a_m such that $d(a_n, a_m) = d_n$, and put $f_n(q) = f_m(q)$ for $q > d_n$, and $f_n(q) = n$ for $q \leq d_n$. If there are a few such points a_m we can take any of them. It is proved in [20] that $f_n(q)$ is well defined and $i : Y \to LV_{\aleph_o}$ is an isometry. Now, to embed Y in LW_2, we choose a countable family $\{Q_n|n \in \mathbf{N}\}$ of pair-wise disjoint dense subsets in $\mathbf{Q}_+$, take characteristic functions χ_n of the sets Q_n , and replace every function f, which equals k on a segment $(s, t]$, by a function which equals χ_k there. More precisely, if the functions $f_0(q), f_1(q), \ldots, f_{n-1}(q)$ are already defined, we put $f_n(q) = f_m(q)$ for $q > d_n$, and $f_n(q) = \chi_n(q)$ for $q \leq d_n$. The specific form of the sets Q_n is inessential. For example, Q_1 may be a set of binary rational numbers $Q_1 = \{m/2^n|m, n \in \mathbf{N}\}$, Q_2 may be a set of ternary rational numbers, $Q_2 = \{m/3^n|m \neq 3^n, m, n \in \mathbf{N}\}, Q_3$ may be a set of 5-nary rational numbers, and so on. Since the Q_n are dense and pair-wise disjoint, it can be proved as in [20] that $\Delta(f_k, f_n) = d(a_k, a_n)$ for any $k < n$, hence $i : Y \to LW_2$ is an isometry. Since LW_2 is complete the closure $[i(Y)]$ in LW_2, contains an isometric image of (X, d). ■

Returning to the case of finite spaces we prove the following analogue to Theorem 6.

Theorem 9. *For any finite ultrametric space (X, d) consisting of $n+1$ points there is a family of at most n two-point spaces $\{(D_k, d_k)|k \leq m\}$, where $m \leq n$, such that (X, d) is isometric to a subspace of their metric product $m\Pi\{(D_k, d_k)|k \leq m\}$.*

Proof. For $n = 0, 1$, and 2 the theorem is obvious. Moreover, here we have an equality $m = n$ instead of the inequality $m \leq n$. Suppose the theorem is proved for all at most n-point spaces. Following the way that we proved Theorems 4 and 5, we take a space $X = \{a_0, a_1, \dots, a_n\}$ enumerated in the same manner and consider the following two cases.

Case 1. $\min\{d(a_i, a_j)|i \neq j\} = d(a_0, a_1) \leq d(a_0, a_2) \leq \cdots \leq d(a_0, a_{n-1}) < d(a_0, a_n) = \max\{d(a_i, a_j)\}$. By the inductive assumption, the set $X_{n-1} = \{a_0, a_1, a_2, \dots, a_{n-1}\}$ can be embedded isometrically in a product $m\Pi\{(D_k, d_k)|k \leq m\}$, where $m \leq n-1$. The point a_n is at the same distance $d(a_0, a_n)$ from any point $a_j (j < n)$, in view of the axiom (Δ). Thus the space X is isometric to a subset of the metric product $m\Pi\{(D_k, d_k)|k \leq m\} \otimes \{0, d(a_0, a_n)\} = m\Pi\{(D_k, d_k)|k \leq m+1\}$, where $m + 1 \leq n$.

Case 2. $\min\{d(a_i, a_j)|i \neq j\} = d(a_0, a_1) \leq \cdots \leq d(a_0, a_{p-1}) < d(a_0, a_p) = \cdots = d(a_0, a_{n-1}) = d(a_0, a_n) = \max\{d(a_i, a_j)\}$. Denote by X_p and X_{p+1} the sets $X_p = \{a_0, a_1, a_2, \dots, a_{p-1}\}$ and $X_{p+1} = \{a_p, a_{p+1}, \dots, a_n\}$. By the inductive assumption, the space X_p can be embedded isometrically in a product $m\Pi\{(D_k, d_k)|k \leq m\}$, where $m \leq p-1$ whereas the space X_{p+1} can be embedded isometrically in a product $m\Pi\{(Di, d_i)|i \leq h\}$, where $h \leq (n-p+1) - 1 = n - p$. As mentioned above all points a_k for $k \geq p$, are at the same distance $d(a_0, a_n)$ from all points a_j for $j < p$ and $d(a_0, a_n) = \max\{d(a_i, a_j)\}$. Therefore (X, d) can be embedded in the product $m\Pi\{(D_k, d_k)|k \leq m\} \otimes m\Pi\{(D_i, d_i)|i \leq h\} \otimes \{0, d(a_0, a_n)\} = m\Pi\{(D_k, d_k)|k \leq m+h+1\}$, where $m + h + 1 \leq n$. ■

Note. The inequality $m \leq n$ cannot be strengthened in general. For any $n \geq 3$, there exist $(n+1)$-point ultrametric spaces that cannot be embedded in $m\Pi\{(D_k, d_k)|k \leq m\}$ for any $m < n$. A few such 4-point spaces are drawn in Figures 2–5 below.

The following particular case of a metric product seems to be the most important for applications in functional analysis. Suppose all factors (X_α, d_α) in a product $m\Pi\{(X_\alpha, d_\alpha)|\alpha \in \mathbf{I}\} = (X_\Pi, d_\Pi)$, are the same space

(X, d) of finite diameter. Then the product $m\Pi\{(X_\alpha, d_\alpha)|\alpha \in \mathbf{I}\} = (X, d)^{\mathbf{I}}$ is none other than a space of all maps from $\mathbf{I}$ to X equipped with a metric of uniform convergence. For an unbounded space (X, d), the pointed product $m\Pi^*\{(X, d, x^*)_\alpha|\alpha \in \mathbf{I}\}$ is a set of all "bounded" maps from I to (X, d) equipped with the same metric. In particular, if (X, d, x^*) is the real line $\mathbf{R}$ with a base point 0, then the pointed product $m\Pi^*\{\mathbf{R}_t|t \in \mathbf{I}\}$ coincides with the set of all bounded real-valued functions $f : \mathbf{I} \to \mathbf{R}$ with the usual norm of uniform convergence $\|f(t)\| = \sup\{|f_{(t)}|,\ t \in \mathbf{I}\}$. Thus an embedding in a product of two-point spaces is at the same time an embedding in Banach space B^τ with the standard sup-norm, $\|x\|_{\sup} = \sup\{|x_t|, t \in \mathbf{I}\}$. This provides us with an interesting characteristic of ultrametric spaces as subsets of Banach spaces.

Criterion 1. Ultrametric spaces are none other than subsets of Banach spaces $B^\tau = (\mathbf{R}^\tau, \|x\|_{\sup})$, whose projections onto any coordinate axis are at most two-point sets.

Recall that by Kuratowski's Theorem [7], any metric space can be embedded in B^τ. Combining the last criterion with Theorem 9 we get the following

Corollary 2. *Any finite ultrametric space (X, d) consisting of $n + 1$ points can be embedded isometrically in the m-dimensional Banach space $B^m = (\mathbf{R}^m, \|x\|_{\sup})$ of dimension $m \le n$ in such a way that projection of the image of X onto any coordinate axis is two-point space.*

It is natural to compare the last corollary and criterion with the well-known theorems on embedding of ultrametric spaces into Euclidean spaces.

Theorem. [14] *Every ultrametric space of weight τ can be isometrically embedded in the generalized Hilbert space H^τ of weight τ.*

Theorem. [14] *Every ultrametric space of cardinality Ψ can be isometrically embedded as a closed subset in the algebraically Ψ-dimensional Euclidean space E^Ψ, but not in E^σ for $\sigma < \Psi$.*

Theorem [14]. *Every finite ultrametric space consisting of $n + 1$ points can be isometrically embedded in the n-dimensional Euclidean space E^n as points in general position. No ultrametric space consisting of $n + 1$ points can be isometrically embedded in E^k for $k < n$.*

In other words $n + 1$ points of an ultrametric space can be considered as vertices of an n-dimensional simplex lying in E^n. The last theorems provide us with another characteristic of ultrametric spaces.

Criterion 2. Ultrametric spaces are none other than the set of vertexes of simplexes lying in Euclidean (Hilbert) spaces, whose two-dimensional faces are isosceles triangles with bases not greater than sides.

There is a close analogy as well as an essential difference between these kinds of embedding. First, a weight of τ-dimensional Hilbert space H^τ is equal to τ whereas a weight of $B^\tau = (\mathbf{R}^\tau, \|x\|_{\sup})$ is $\aleph_0^\tau$ ($= 2^\tau$ for infinite τ). Further the smallest dimension of the Euclidean space, which contains a given $(n+1)$-point ultrametric space X is n for all $(n+1)$-point ultrametric spaces. And the smallest dimension of the similar Banach space $B^m = (\mathbf{R}^m, \|x\|_{\sup})$ containing a given $(n+1)$-point ultrametric space (X, d), depends on the metric d of a space X. It can be smaller than n, more precisely, $[\log_2(n+1)] \leq m \leq n$. The difference arises whenever $n \geq 3$. Let us illustrate it for the smallest nontrivial value, $n = 3$. In view of the last theorem, any four-point ultrametric space can be viewed as vertexes of tetrahedron $\subset E^3$ of one of the following six types.

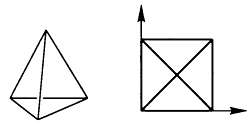


Figure 1

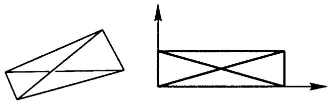


Figure 6

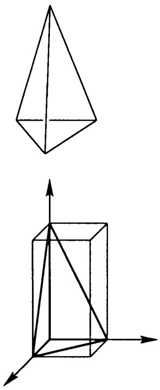


Figure 2

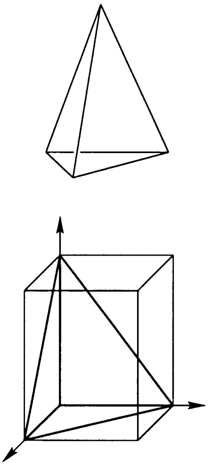


Figure 3

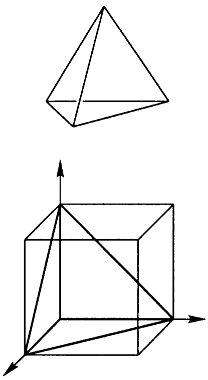


Figure 4

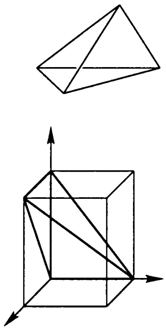


Figure 5

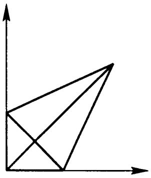


Figure 2a

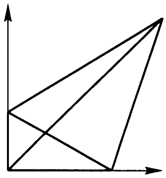


Figure 3a

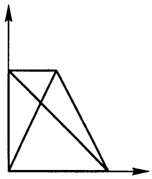


Figure 4a

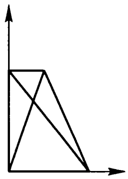


Figure 5a

The set $V = \{d(x, y) | x \neq y \in X\}$ of non-zero values of the metric d has cardinality $|V| = 1, 2, 3, 2, 3, 2$ in each of these cases respectively. In case 1 (a regular tetrahedron), there is an obvious possibility to embed X in **two**-dimensional Banach space B^2, as vertexes of a square, see Figure 1. Here all segments, both sides and diagonals, have the same Banach length, $|V| = 1$. The same possibility exists in case 6, although $|V| = 2$ there (see

Figure 6). In cases 3 and 5, $|V| = 3$ therefore it is impossible to embed X in a product of two two-point spaces. It is impossible to find such an embedding in cases 2 and 4 either. However, there exist embeddings of all these types of simplexes in two-dimensional Banach space with three-point projections on coordinate axes (see Figures 2a–5a).

Embedding of ultrametric spaces into Euclidean spaces enables us to apply the theory of ultrametric spaces to linear and convex programming (to the simplex method, in particular). We may hope that the theorems on embedding in Banach spaces will also find appropriate applications. We complete this section by the following problems.

Problem 3. Is it possible to embed an arbitrary ultrametric space into an arbitrary Banach space (a Banach space equipped with an arbitrary norm, not necessarily the sup-norm) of appropriate weight, in particular, of the same weight?

Problem 4. Does there exist, for any natural n, a number $N(n)$ such that any n-point ultrametric space can be embedded isometrically in every $N(n)$-dimensional Banach space?

As mentioned above, $N(n) = n - 1$ for Euclidean spaces [14]. Among general Banach spaces, the spaces L_p of Lebesgue integrable functions on $\mathbf{R}$ are the most important. For these spaces, the embedding problem was stated by Prof. Sergey Nikolski in the beginning of the 1970s [30].

Problem 5 [S. Nikolski, 30]. Is it possible to embed any separable ultrametric space in the space L_p for any $p \geq 1$.

Prof. Israel Gelfand has recently mentioned that for $p = 1$ the affirmative answer to Nikolskiy's problem follows from the next theorem.

Theorem [18]. *Every separable ultrametric space can be embedded isometrically in the Lebesgue space $L(\mathbf{R})$.*

Here $L(\mathbf{R})$ denotes a space of Lebesgue measurable subsets of $\mathbf{R}$ with the metric being equal to the measure of symmetrical difference $d(A, B) = \mu(A \Delta B)$.

Corollary (I.Gelfand). *Every separable ultrametric space can be embedded isometrically in the space $L_1(\mathbf{R})$ of Lebesgue integrable functions on $\mathbf{R}$ with the norm $\|f(x)\|_1 = \int |f(x)| dx$.*

To prove this, it is enough to assign a characteristic function $\chi_A(x)$ to any measurable subset $A \subset \mathbf{R}$ and note that $\|\chi_A(x) - \chi_B(x)\|_1 = \mu(A \Delta B)$. For $p = 2, L_2(\mathbf{R})$ is none other than Hilbert space H, so the problem is also solved in the affirmative by Theorem [14] adduced above. In 1975 A. Timan gave a very partial answer to Nikolski's problem for a certain type of countable spaces [29]. However, in general, the problem is still open.

4 Reflectivity and Scanning Programs

Lemmas 1 and 2 as well as others that concern limits of direct and inverse spectra, pull-backs and push-outs, equalizers and co-equalizers, and other categorical operations show that the standard categorical procedures being applied to ultrametric spaces give us again ultrametric spaces [11, 15]. Moreover, we always obtain ultrametric spaces as the result of the action on these spaces of basic metric functors. These are: a trimming functor **METR** $\to$ **METR$_c$**; a completion functor **METR** $\to$ Complete-**METR**; an orbital functor (a functor of passing to invariant metric $(X, d) \to (X, d_G)$ and to a space of orbits $X/_G$) for spaces acted on by a compact group G; the Hausdorff exponential functor $(X, d) \to (\mathrm{Hexp}X, d_H)$, where Hexp X is the space of all (bounded) closed subsets of X with the Hausdorff metric $d_H(A, B) = \inf\{\epsilon | A \subset O_\epsilon(B), B \subset O_\epsilon(A)\}$; functors of passing to various functional spaces with metric of uniform convergence (see Section 3 above), and so on [11, 12, 17]. These can be resumed as a non-formal principle: **'Ultrametrics generate Ultrametrics'**. A reader would probably think that ultrametric spaces form an absolutely closed class with no relations to other spaces. Fortunately, this is not so. Non-expanding (uniform, continuous) maps connect ultrametric spaces with all other metric (uniform, topological) spaces. The latter are at the same time images and pre-images of ultrametric spaces under "nice" maps.

Theorem 10. *For every cardinal τ, there is a complete ultrametric space $L^*(\tau)$ such that any metric space of weight $\leq \tau$ is an image of $L^*(\tau)$ under a non-expanding open map.*

In other words $L^*(\tau)$ is a universal pre-image of all metric spaces of weight $\leq \tau$ (*the initial object*). This theorem generalizes a theorem due to Holsztynski [5], who studied initial objects in certain subcategories of **METR** (without the requirement for a map to be *open*, and under fairly strong additional assumptions, which imply, in particular, common boundedness of weights, cardinalities, and diameters of the considered spaces). Note that for all spaces of diameter at most c and cardinality at most σ, the theorem is trivial without the requirement 'f is open'; a discrete space of cardinality σ and pair-wise distances $= c$, is obviously a universal inverse image. Theorem 10 can be easily deduced from the next one.

Theorem [13]. *For every cardinal τ, there is a complete ultrametric space L_τ of weight τ such that any metric space (X, d) of weight $\leq \tau$ is an image of a subspace $L(X) \subset L_\tau$ under a non-expanding open map $f : L(X) \to X$ with compact pre-images of points and totally bonded pre-images of compact subsets $K \subset X$.*

Proof. The space L_τ is naturally defined. It is a set of infinite (in both directions) sequences $\mathbf{a} = (\dots, 0, \alpha_{-m}, \alpha_{-m+1}, \dots, \alpha_{-1}, \alpha_0, \alpha_1, \dots, \alpha_n, \dots)$ of ordinals $\alpha_n < \omega(\tau)$ containing only finitely many non-zero terms with negative subscript. A metric $d(\mathbf{a}, \mathbf{b})$ on L_τ is defined in a Baire manner $d(\mathbf{a}, \mathbf{b}) = 2^{n-1}$, where $n = \min\{k | \alpha_k \neq \beta_k\}$. For the rest of a proof see [13]. ■

The space L_τ generalizes the Baire space B_τ in the same way as a Laurent series generalizes a Taylor series in complex analysis. That is why we denote it by L_τ. We can even replace L_τ by B_τ if we reduce the requirement 'f is non-expanding' to 'f is continuous'. This theorem contains a few well known theorems. E.g., requirement 'f is continuous, open, and compact' gives us Nagami's Theorem [23]. Omitting the third assertion we get Morita's Theorem [21] (proved by Hausdorff [4] for separable spaces with B_τ replaced by $B_{\aleph_o}$). Morita's Theorem was generalized to all T_0-spaces of countable character by V. Ponomarev [24]. Note that generally the map f cannot be made perfect (pre-images of compact subsets need not be compact) because perfect maps do not augment dimension, whereas all ultrametric spaces are zero-dimensional.

To prove Theorem 10 we choose a sequence $\mathbf{0} = (\dots, 0_{-m}, \dots 0_{-1}, 0_0, 0_1, \dots, 0_n, \dots)$ a base point in L_τ, choose any point x^* in X and verify that, by definition [13], $L(X)$ contains $\mathbf{0}$, and the map $f : L(X) \to X$ takes $\mathbf{0}$ to x^*. Then the space $L^*(\tau)$ can be defined as a pointed product of all pair-wise non-isometric subsets $L(X) \subset L_\tau$ containing $\mathbf{0}$. The following proposition completes the proof.

Proposition 11. *Canonic projections $p_\beta : m\Pi\{(X_\alpha, d_\alpha, x^*_\alpha)|\alpha \in I\} \to (X_\beta, d_\beta, x^*_\beta)$ of a pointed product onto any factor $(X_\beta, d_\beta, x^*_\beta)$ are open.*

Proof. Let G be an open subset of $m\Pi^*\{(X_\alpha, d_\alpha, x^*_\alpha)|\alpha \in \mathbf{I}\}$. For any $\{x_\alpha\}$ in G there is a positive ϵ such that $O_\epsilon(\{x_\alpha\}) \subset G$. Let $O_\epsilon(x_\beta)$ be an ϵ-neighborhood $O_\epsilon(x_\beta)$ of the point $x_\beta = p_\beta(\{x_\alpha\})$ in X_β and $y_\beta \in O_\epsilon(x_\beta)$. Then the set of all $\{z_\alpha\}$ such that $z_\alpha = x_\alpha \forall \alpha \neq \beta, z_\beta = y_\beta \in O_\epsilon(x_\beta)$, is contained in $O_\epsilon(\{x_\alpha\})$. Hence $p_\beta(O_\epsilon(\{x_\alpha\})) = O_\epsilon(x_\beta)$. Therefore, $p_\beta(G)$ is open. ■

Problem 6. Does there exist a universal ultrametric pre-image of weight τ for all metric spaces of weight at most τ under non-expanding open mappings?

On the other hand ultrametric spaces are not only pre-images but also images of general metric spaces under non-expanding maps. Moreover, for any metric space (X, d), there exists a greatest element (uX, d_u) in the set of all ultrametric images of X under non-expanding maps. This is called *an ultrametrization* of X. The word "greatest" is explained in the following theorem.

Theorem 11. *For every metric space (X, d), there are an ultrametric space (uX, d_u) and a non-expanding surjection $u : (X, d) \to (uX, d_u)$ such that for any non-expanding map $f : (X, d) \to (Y, r)$ from X to ar-*

bitrary ultrametric spaces (Y, r) there exists a unique non-expanding map $uf : (uX, d_u) \to (Y, r)$ such that $uf \cdot u = f$.

Proof. To prove the theorem we recall the notions of ϵ-chain, ϵ-connectedness, and Cantor connectedness (due to G. Cantor and F. Hausdorff) and describe their behavior under non-expanding maps. Let (X, d) be a metric space and $a, b \in X$. A sequence $a = x_0, x_1, \dots, x_{n-1}, x_n = b$ is called *an ϵ-chain* between a and b provided $d(x_{k-1}, x_k) \le \epsilon$ for any $k \le n$. Two points a and b are called *ϵ-linked* if there is an ϵ-chain between them. They are called *linked* if they are ϵ-linked for any $\epsilon > 0$. A space (X, d) is *Cantor connected (linked)* if any two points in X are linked [10]. It is easily verified that the binary relation '$x \sim y$ iff x and y are linked', is an equivalence relation. Denote by $[x]$, uX, and $u : X \to uX$ the equivalence class of a point x, the quotient space $X/\sim$, and the natural projection $u : X \to X/\sim$, respectively. To introduce a metric d_u on uX, we put $d_u(x, y) = \inf\{\epsilon | x$ and y are ϵ-linked $\}$. It is obvious that $d_u(x, y) \ge 0, d_u(x, x) = 0, d_u(x, y) = d_u(y, x)$. Next, if x and y are ϵ-linked, and y and z are δ-linked, then x and z are $\max[\epsilon, \delta]$-linked. This implies that $d_u(x, y)$ satisfies the ultrametric Axiom (Δ). Thus $d_u(x, y)$ is a pseudo-ultrametric on X. Clearly $d_u(x, y) = 0$ iff x and y are linked, hence $d_u(x, y)$ is well defined on the quotient space and it is an ultrametric there, i.e., it satisfies the axiom '$d([x], [y]) = 0$ implies $[x] = [y]$'. Since any pair of points x and y is a $d(x, y)$-chain between them, $d_u(x, y) \le d(x, y)$. Therefore the natural projection $u : (X, d) \to (uX, d_u)$ is non-expanding. ■

Lemma 3. *No two points a and b in an ultrametric space (Y, r) are ϵ-linked for any $\epsilon < r(a, b)$.*

Proof. For the chain $a = x_0, x_1, x_2 = b$ consisting of three points, this follows from the Axiom Δ (= the isosceles property). The general case can be proved by induction over a length of chain. ■

This property completely characterizes ultrametric spaces (among general metric spaces) because of the following

Lemma 4. *A metric space (X, d) is ultrametric if and only if no two points a and b in X are ϵ-linked for any $\epsilon < d(a, b)$.*

Proposition 12. *Any non-expanding map takes an ϵ-chain to an ϵ-chain.*

Corollary 3. *Non-expanding maps preserve Cantor connectedness.*

Let (Y, r) be an ultrametric space and $f : X \to Y$ be a non-expanding map. By Lemma 3, for any two points x and y in X, the images $f(x)$ and $f(y)$ are not ϵ-linked for any $\epsilon < r(f(x), f(y))$. In view of Proposition 12, x and y are not ϵ-linked for any $\epsilon < r(f(x), f(y))$. Hence $d_u(x, y) \ge r(f(x), f(y))$. This implies, first, that for $x \sim y, f(x) = f(y)$, i.e., the map $uf : uX \to Y$, defined as $uf([x]) = f(x)$, is well-defined and completes the diagram below. Second, it is non-expanding in view of the inequality $d_u([x],[y]) = d_u(x,y) \geq r(f(x),f(y)) = r(uf([x]),uf([y]))$.

$$\begin{tikzcd}
(uX,d_u) \arrow[r, "uf"] & (Y,r) \\
(X,d) \arrow[u, "u"] \arrow[ur, "f"'] &
\end{tikzcd}$$

Corollary 4. *The subcategory* **ULTRAMETR** *is a reflective subcategory in* **METR**.

Proof. For any non-expanding map $f : (X,d) \to (Z,d')$ of general metric spaces, a composition $u \cdot f : X \to Z \to uZ$ is a non-expanding map from X to an ultrametric space uZ. Hence it can be lifted to a map $uf : uX \to uZ$ such that the following diagram commutes

$$\begin{tikzcd}
(uX,d_u) \arrow[r, "uf"] & (uZ,d'_u) \\
(X,d) \arrow[u, "u"] \arrow[r, "f"] & (Z,d') \arrow[u, "u"']
\end{tikzcd}$$

The properties $u(fg) = uf \cdot ug$ and $u1_X = 1_{uX}$ are obvious. Thus u is a covariant reflective functor from **METR** to **ULTRAMETR**. ■

Example. Let $X = X_1 \cup X_2 = [-1,0) \cup \{1/2^n | n \geq 0\}$ be a subset of the real line **R** with the usual metric. Then the space $uX = \{0\} \cup \{1/2^n | n \geq 1\}$ is a subset of $(\mathbf{R}_+, \max)$ with the max-metric. $X_1 = [-1,0)$ is open in X whereas $u(X_1) = \{0\}$ is not open in uX, X_2 is closed in X whereas $u(X_2)$ is not closed in uX. Thus the reflective map u is neither open, nor closed in general.

Definition. Two metric spaces X and Z are said to be *ultrametric equivalent* (or *u-equivalent*) if their ultrametrizations are isometric, $uX = uZ$.

It follows from Corollary 3 that all Cantor connected spaces (and only they) are u-equivalent to a singleton.

Theorem 12. Any metric space consisting of n points is u-equivalent to an n-point subset of the real line.

Proof. Since any metric space is u-equivalent to its own ultrametrization, it is enough to prove the theorem for ultrametric spaces only. An ultrametrization of an n-point metric space is n-point. For $n = 1$ and 2 the assertion of the theorem is trivial. A three-point ultrametric space $X = \{a_0, a_1, a_2\}$ with $d(a_0, a_1) = d_1 \leq d(a_0, a_2) = d(a_0, a_2) = d_2$, is u-equivalent to three points $0, d_1$ and $d_1 + d_2$ in $\mathbf{R}$. Suppose, for any k-point ultrametric space $X_k = \{a_0, a_1, \ldots, a_{k-1}\}$, for $k \leq n$, there is a space $Z_k = \{b_0, b_1, \ldots, b_{k-1}\} \subset \mathbf{R}$ with $b_0 < b_1 < \cdots < b_{k-1}$, which is u-equivalent to X_k. Using Lemma 3 [14] for the last time throughout the paper, we see that, in case 1, it is enough to add a point $b_n = b_{n-1} + d(a_0, a_n)$ to the space Z_n to get a u-equivalent space $Z_{n+1} = Z_n \cup \{b_n\}, uZ_{n+1} = uX (= X)$. In case 2, we take the spaces $Z_p = \{b_0, b_1, b_2, \ldots b_{p-1}\}$ and $Z_{p+1} = \{b_p, b_{p+1}, \ldots b_n\}$, which are u-equivalent to X_p and X_{p+1} respectively, and move Z_{p+1} as a rigid set along the real line in such a way that $b_p - b_{p-1} = d(a_0, a_p) = \cdots = d(a_0, a_n)$. Location of the set $Z = Z_p \cup Z_{p+1}$ on the real line implies that $uZ = X$. ■

Using this theorem E. V. Schepin created an effective scanning algorithm and realized it on an IBM-compatible computer [28]. Other applications of the last theorem to the problem of pattern recognition will be published in future.

Acknowledgment. The author is grateful to Austin Melton, George Strecker and Horst Herrlich for useful discussions.

References

[1] P.S. Alexandroff. On the theory of topological spaces, *Soviet Math. Dokl*, **2** (1936), 51–54 (in Russian).

[2] B. Flagg and R. Kopperman. Computational Models for Ultrametric Spaces, *XIII Conference on Mathematical Foundations of Programming Semantics*, 1997, 83–92.

[3] J. de Groot. Non-Archimedean metrics in topology, *Proc. A.M.S.*, **7:6** (1956), 948–956. MR 18-325.

[4] F. Hausdorff. Uber innere Abbildungen, *Fund. Math.* **23** (1934), 279-291.

[5] W.Holsztynski. Initial and universal metric spaces, *Proc. A.M.S.*, **58** (1976), 306–310. MR 56 #13174.

[6] M. Krasner. Nombres semi-réels et espaces ultramétriques, *C. R. Acad. Sci.*, Paris, **219** (1944), 433–435. MR 7-364.

[7] K. Kuratowski. Quelques problèmes concernant les espaces métriques non-sépaprable, *Fund. Math.*, **25** (1935).

[8] J.D. Lawson. Spaces of maximal points, preprint, 1995.

[9] A.J. Lemin. On isosceles metric spaces, *Functional Analysis and its Applications*, Moscow, MSU-Press, 1884, 26–31 (in Russian).

[10] A.J. Lemin. Proximity on isosceles spaces, *Russian Math. Surveys*, **39:1** (1984), 143–144. MR 86j: 54059.

[11] A.J. Lemin. On the stability of the property of a space being isosceles, *Russian Math. Surveys*, **39:5** (1984), 283–284. MR 86b: 54029.

[12] A.J. Lemin. Transition functors to a function space in the uniform topology, *Russian Math. Surveys*, **40:6** (1985), 133–134.

[13] A.J. Lemin. Inverse images of metric spaces under non-expanding open mappings, *Russian Math. Surveys*, (1988), 214–215.

[14] A.J. Lemin. Isometric embedding of isosceles (non-Archimedean) spaces in Euclidean spaces, *Soviet Math. Dokl.*, **32:3** (1985), 740–744. MR 87h: 54056.

[15] A.J. Lemin. An application of the theory of isosceles (ultrametric) spaces to the Trnkova-Vinarek theorem, *Comment. Math. Univ. Carolinae*, **29:3** (1988), 427–434.

[16] A.J. Lemin. Spectral decomposition of ultrametric spaces and topos theory, *14^{th} Summer Conf. on Topology & App.*, Abstracts: Long Island Univ., Aug. 4-7, 1999, 22.

[17] A.J. Lemin. The category of ultrametric spaces is isomorphic to the category of complete, atomic, tree-like, real graduated lattices **LAT***, to appear.

[18] A.J. Lemin. Isometric embedding of ultrametric (=non-Archimedean) spaces in the Lebesgue spaces, to appear.

[19] A.J. Lemin and V.A. Lemin. On uniform rationalization of ultrametrics, *Topology Proceedings*, **22**, Summer 1997, 275–283.

[20] A.J. Lemin and V.A. Lemin. On a universal ultrametric space, *Topology and its Applications*, to appear.

[21] K. Morita. Normal families and dimension theory for metric spaces, *Math Ann.*, **128** (1954), 350–362. MR 16-501.

[22] S. Mrowka. Recent results on E-compact spaces, *Proc. 2^{nd} Intl. Conf. General Topology and its Applications*, 1972. MR 50-14673.

[23] K. Nagami. A note on Hausdorff spaces with star-finite property. I, II, *Proc. Japan Acad.*, **37** (1961), 131–134, 189–192. MR 26 #1852.

[24] V.I. Ponomarev. Axioms of countability and continuous mappings, *Bull. Acad. Pol. Sci.*, **8:3** (1960), 127–134. MR 22 #7109.

[25] P. Roy. Non-equality of dimensions for metric spaces, *Trans. A.M.S.*, **134** (1968), 117–132. MR 37 #3544.

[26] Yu.M. Smirnov, On proximity spaces, *Math. Sbornik*, **31** (1952), 543–574 (in Russian), A.M.S. Trans. Ser. 2, **38**, 5–35.

[27] Yu.M. Smirnov, On the dimension of proximity spaces, *Math. Sbornik*, **38** (1956), 283–302 (in Russian), A.M.S. Trans. Ser. 2, **38**, 37–73.

[28] E.V. Schepin. Private communication.

[29] A.F. Timan. On isometric embedding of certain countable ultrametric spaces in the L_p-space, in [30], 314–326 (in Russian).

[30] Trudy of Mathematical Institute of the USSR Acad. Sci., **134** (1975). Collection of articles dedicated to Sergey Mihailovich Nikolski on the occasion of his seventieth birthday, Moscow, 1975.

[31] J.E. Vaughan. Universal ultrametric spaces of smallest weight, 1999, 1–7, preprint.

[32] J.E. Vaughan. Universal ultrametric spaces of smallest weight, *14th Summer Conf. on Topology & App.*, Abstracts: Long Island Univ., Aug. 4-7, 1999, 38–39.

Vladimir A. Lemin
Department of Mathematics
Pennsylvania State University
224 McAllister Building
University Park, PA 16802-6401
lemin@math.psu.edu

The Copnumber of a Graph is Bounded by $\left\lfloor \frac{3}{2} \text{ genus } (G) \right\rfloor + 3$

Bernd S. W. Schroeder

ABSTRACT We prove that the copnumber of a finite connected graph of genus g is bounded by $\lfloor \frac{3}{2}g \rfloor + 3$. In particular this means that the copnumber of a toroidal graph is bounded by 4. We also sketch a proof that the copnumber of a graph of genus 2 is bounded by 5.

Key words: Copnumber, genus, retraction, toroidal graph.

AMS Subject Classifications (2000): 05C10, 05C75, 57M15.

1 Introduction

The game of cops and robber on a graph was simultaneously conceived by Nowakowski and Winkler and by Quilliot for one cop and one robber, cf. [13] and [15]. Subsequently, the game was generalized to several cops and one robber by Aigner and Fromme, cf. [1]. It has since drawn considerable attention as the references [2]–[8], [10], [11] and [16] evidence. A comprehensive list of references can be found in [11].

In this game, player $\mathbb{C}$ controls the cops and player $\mathbb{R}$ controls the robber. Cops and robber are located at the vertices of the graph and move along the edges. First, player $\mathbb{C}$ chooses vertices to place the cops, then player $\mathbb{R}$ chooses a vertex for the robber. The players then move alternately, starting with $\mathbb{C}$. Throughout the game each player knows the location of the opposition. At each move, player $\mathbb{R}$ chooses a (possibly empty) subset S of the cops to be moved and moves each cop in S to a vertex adjacent to its current position. Player $\mathbb{R}$ is allowed to leave the robber stationary or move it to a vertex adjacent to its current position. Player $\mathbb{C}$ wins if one cop occupies the same vertex as the robber after some move. If player $\mathbb{R}$ can avoid this situation indefinitely, player $\mathbb{R}$ wins. The *copnumber* of a graph is the smallest number of cops that is needed to guarantee player $\mathbb{C}$ a winning strategy (hereafter called a *capture strategy*).

For example, on a path of length n, one can start one cop at one end of the path and march it across to the other end. In this fashion, independent of the robber's actions, the cop will capture the robber at the latest by the time it arrives at the opposite end of the path. Thus the copnumber for a path is 1. On the other hand, on a cycle of length ≥ 4 the robber can evade one cop indefinitely. Initially place the robber as far away from the cop as possible. Then always move in the direction that maximizes the distance to the cop. Thus the copnumber of a cycle with ≥ 4 vertices is not 1. In fact, it is 2. Start two cops at the same vertex, one moves clockwise, the other counterclockwise.

Upper bounds on the copnumber have been found for special classes of graphs such as graphs with excluded minors (cf. [3], Theorems 1, 2 and 3), graphs that have certain types of retractions (cf. [5], Theorems 3.2 and 3.5), and products of trees (cf. [10], Theorem 4). Lower bounds on the copnumber are given in [1], Theorems 3 and 5 and [7], Theorems 1.1, 1.3 and 1.5. The complexity of determining if the copnumber is bounded by k is shown to be polynomial in [5], Theorem 2.1, while the complexity of determining the copnumber is shown to be EXPTIME complete in [8], Theorems 3 and 4. Bridged graphs are characterized using graphs of copnumber 1 (i.e., dismantlable graphs) in [4], Corollary 2.6 and variants of the game are considered in [11]. A new geometric approach using the strong isometric dimension of the graph can be found in [6].

The most visual and possibly most attractive results are upper bounds on the copnumber using graphical parameters such as the genus and the crosscap number (cf. [1], Theorem 6, [3], Corollary 3, [16], Theorem 1). The genus of a graph G is the smallest number g such that G has a planar drawing on a surface of genus g. For an introduction to topological graph theory cf. [17], for facts on surfaces, cf. [9]. It was shown by Aigner and Fromme in [1], Theorem 6 that the copnumber of planar (genus 0) graphs is bounded by 3. Subsequently, Quilliot showed in [16], Theorem 1 that the copnumber of a graph of genus g is bounded by $2g+3$. While examples on p.5 in [1] showed that there are planar graphs whose copnumber is indeed 3, no examples of graphs of genus $g \geq 1$ with copnumber $2g+3$ are known and the last question on p. 92 in [16] asks if such graphs exist at all. In the later paper [3] Andreae asks more specifically if the copnumber for toroidal graphs is bounded by 4 or even 3 (cf. [3], open problem 3, p.46).

In this paper (cf. Theorem 3.3) we show that the copnumber of a graph of genus g can be bounded by $\left\lfloor \frac{3}{2}g \right\rfloor + 3$. In particular, this implies that the copnumber of a toroidal graph is bounded by 4. While this result answers the question in [16] and (at least partially) the open problem 3 in [3], it opens up the analogous questions for the new bounds given here. The author conjectures that the copnumber of a graph of genus g can at least be bounded by $g+3$ and that the copnumber of toroidal graphs is in fact bounded by 3. In section 4 we present some insights that show that new ideas are needed if a bound of $g+3$ on the copnumber is to be proved, since the obvious generalizations of the known approaches (including the one presented here) cannot work. As further backup for the $g+3$ conjecture we present in Section 5 the sketch of a proof that the copnumber of a graph of genus 2 is bounded by 5.

In a future paper (cf. [12]) R. Nowakowski and the author will improve the upper bound on the copnumber in terms of the crosscap number given in [3], Corollary 3 (as is suggested there in open question 2).

2 Basic Definitions and Preliminary Results

Throughout, the robber, respectively its current position, will be denoted by ρ. The cops, respectively their current positions, will be denoted by $\gamma_1, \gamma_2, \ldots$. Induced subgraphs of G on a vertex set W will be denoted $G[W]$. The capture strategies outlined in [1] and [16] have the same underlying theme. Cops are placed and move according to a strategy that limits the movement of the robber to a certain subgraph of the original graph.

Definition 2.1 Let $G = (V, E)$ be a graph and let $P \subseteq V$. We will say that a cop γ *has been moved to block* P iff cop γ moves on P according to a strategy that makes it impossible for the robber to reach P without being captured by γ. If cop γ had been moved to block P, we will say γ is *free to move away from* P iff other cops have been moved into positions and follow strategies that do not allow the robber to reach P without being captured by a cop other than γ. A set of cops moved to block a set B of vertices is defined analogously.

Definition 2.2 Let $G = (V, E)$ be a finite graph with k cops, moved to block a set B of vertices, and a robber on it. We define the *robber territory* (cf. [1]) to be the set of vertices R that the robber can reach without being captured by any of the cops blocking B.

A *retraction* on $G = (V, E)$ is a map $r : V \to V$ such that $r^2 = r$ and such that $\{x, y\} \in E$ implies $r(x) = r(y)$ or $\{r(x), r(y)\} \in E$. The set $r[V]$ is also called a *retract* of G. If $r[V]$ is a retract, then one cop γ_{k_0}, placed at $r(\rho)$ and moving such that after each cop move it is at $r(\rho)$, prevents the robber from entering $r[V]$. Indeed, if it ever was the case that $\rho = r(\rho)$, then γ_{k_0} would capture the robber in the next move. Thus to move to block a retract $r[V]$, it is enough to move sufficiently many cops into $r[V]$, then run a capture strategy in $G[r[V]]$ to capture $r(\rho)$ and subsequently let one cop "shadow" the robber (cf. [6], after Lemma 2.1) as described above.

An isometric (also referred to as geodesic) path $P = \{p_0, \ldots, p_n\}$ is a retract via $r(v) := p_{\min\{d(v,p_0),n\}}$ with $d(x, y)$ being the distance between x and y in G. Since paths have copnumber 1, one cop is thus good enough to move a cop to block an isometric path. Retractions are the most common way to move a cop to block a set of vertices. Note however that a capture strategy only needs to take into account the robber territory R and the set of vertices B that are adjacent to vertices of R. Thus retractions we use need not be retractions on G, but only on an appropriate subgraph of G.

Naturally limitations to the movement of the robber do not translate into limitations for the movement of the cops.

As a capture strategy proceeds, the robber territory generally shrinks (eventually to an empty set), i.e., the robber territory will change as a strategy progresses. In the strategy in [16] g pairs of two cops are blocking g non-null-homotopic cycles to limit the robber's moves to a planar graph. The key realization is that a minimal non-null-homotopic cycle is made up by two isometric paths that can be blocked with a cop each. Since afterwards the robber cannot cross the cycle, its movements are now restricted to a graph with genus at least one less than the original graph. (One of the handles of the surface on which the graph is drawn has been "cut".) Repeat this procedure until the robber territory is planar.

The strategy in [1] for planar graphs successively shrinks the robber territory. Two cops patrol on the two "halves" of a cycle that "surrounds" the robber territory, while the third cop is moved to a path across the robber territory to produce the next partition of the robber territory. The partitions are done in such a way that after the third cop has been assigned a strategy, one of the first two cops is free to move away from its previous assignment on the boundary of the previous robber territory. This "recycled" cop is then used to produce the next partition of the robber territory. Our strategy lifts this underlying idea to graphs of higher genus, merging it with Quilliot's idea of limiting the robber's moves to graphs of smaller and smaller genus.

The next lemma, though technical, is very useful in reducing the number of special cases to be considered when constructing a capture strategy on a non-planar graph. Essentially it connects capture strategies on a given graph to capture strategies on quotients of the graph.

Definition 2.3 Let $G = (V, E)$ and $G' = (V', E')$ be graphs. Then G is called a **quotient** of G' iff there is a surjective map $F : V' \to V$ (called the **quotient map**) such that:

1. $\{x, y\} \in E'$ implies $F(x) = F(y)$ or $\{F(x), F(y)\} \in E$, (i.e., F is a homomorphism).

2. For every edge $\{a, b\} \in E$ there is an edge $\{x, y\} \in E'$ such that $\{F(x), F(y)\} = \{a, b\}$.

Lemma 2.4 *Let $G = (V, E)$ be a graph with cops and a robber on it. Suppose*

1. *There is a set C_1 of cops that have been moved to block the set B of vertices,*

2. *R is the set of the vertices of V which ρ can reach without being captured by any cop in C_1,*

3. *$G' = (V', E')$ is a graph such that $G[B \cup R]$ is a quotient of G',*

4. *The restriction of the quotient map $F : V' \to B \cup R$ to $F^{-1}[R]$ is an isomorphism between $F^{-1}[R]$ and R,*

5. *There is a capture strategy in G' that uses $a + k$ cops and has at least a of the cops moving to prevent $F^{-1}(\rho)$ from reaching $F^{-1}[B]$ (i.e., limiting $F^{-1}(\rho)$'s moves to $G'[F^{-1}[R]]$).*

Then there is a capture strategy on G that uses at most $|C_1| + k$ cops. If we replace 5 with

5'. *There is a strategy in G' with $a + k$ cops, which has at least a of the cops moving to prevent $F^{-1}(\rho)$ from reaching $F^{-1}[B]$ (i.e., limiting $F^{-1}(\rho)$'s moves to $G'[F^{-1}[R]]$) and which eventually moves cops to block a set B' of vertices, such that $F^{-1}(\rho)$ cannot reach any vertex of $F^{-1}[B]$ without being captured by one of the cops that were moved to block B',*

Then there is a strategy on G that uses at most $|C_1| + k$ cops, which eventually moves cops to block the set of vertices $F[B']$, such that the robber cannot reach any vertex of B without being captured by one of the cops that were moved to block $F[B']$.

Proof. The cops in C_1 have been moved to prevent ρ from reaching B and a cops in G' have been moved to prevent $F^{-1}(\rho)$ from reaching $F^{-1}[B]$. Let $\gamma_1', \ldots, \gamma_k'$ be the cops on G' that do not block $F^{-1}[B]$. Move $\gamma_1', \ldots, \gamma_k'$ according to the strategy in 5, or 5' and move γ_i in G to $F(\gamma_i')$ in every cop move.

If 5 holds, ρ is captured in G when $F^{-1}(\rho)$ is captured in G'. If 5' holds, ρ is blocked from $F[B']$ once $F^{-1}(\rho)$ is blocked from B'. Indeed, ρ cannot reach B without capture in $F[B']$, since otherwise $F^{-1}(\rho)$ could reach $F^{-1}[B]$ without capture on B'. ∎

We will use Lemma 2.4 to formulate readable capture and limitation strategies and sub-strategies rather than accumulating more and more hypotheses on one existing strategy.

We also formally define what a drawing of a graph on a surface is. In our proofs we will use descriptive language when it comes to referring to drawings and also to modifying drawings to obtain drawings of new graphs, trusting the reader will be able to fill in the necessary (tedious) parametrizations, etc.

Definition 2.5 Let $G = (V, E)$ be a finite graph, let S be a 2-dimensional surface and let $A(S)$ be the set of all continuous, injective arcs $a : [0, 1] \to S$. A *drawing* of G on S (or a *realization* of G on S) is a map $d : V \cup E \to S \cup A(S)$ such that

1. $d[V] \subseteq S$,

2. $d[E] \subseteq A(S)$,
3. For all $\{x,y\} \in E$ we have $d(x) = d(\{x,y\})(0)$ and $d(y) = d(\{x,y\})(1)$ or $d(x) = d(\{x,y\})(1)$ and $d(y) = d(\{x,y\})(0)$,
4. For distinct edges e_1 and e_2 and all $t_1, t_2 \in (0,1)$ we have $d(e_1)(t_1) \neq d(e_2)(t_2)$.

Notation 2.6 To abbreviate notation we will give paths in a graph just as a list of vertices with the understanding that the path starts at the first vertex of the list, ends at the last one and consecutively listed vertices are adjacent. We indicate cycles in similar fashion. Paths and cycles indicated in this fashion need not be induced paths or cycles and can have repeated vertices. If a cycle with vertices C or a path with vertices P is to be an induced cycle/path, we will indicate this by referring to $G[C]$ or $G[P]$ as the cycle/path.

Also, if d is a drawing of G and H is a subgraph of G, we will write $d[H]$ for $d|_{V(H)\cup E(H)}$, the induced drawing of H. Moreover, as this will not cause any confusion, we will not distinguish between the drawing and the set of points in S which are the images of the vertices or which are in the images of the arcs.

Finally, ε will denote a "small enough" positive number which will not be specified any further.

Definition 2.7 We will call the plane $\mathbf{R}^2$ with g handles attached inside the unit square $\{(x,y) : |x| < 1, |y| < 1\}$ the *g-unit-handled plane.* A drawing of a graph on a g-unit-handled plane is called *good* iff no drawing of any vertex or edge enters $\{(x,y) : |x| > 1 \text{ or } |y| > 1\}$.

Definition 2.8 Let $G = (V, E)$ be a finite graph, let S be a 2-dimensional surface and let d be a drawing of G on S. Suppose $v \in V$ and U is an open set in S such that $d(v)$ is on the boundary δU of U. Then the edge $\{v,w\}$ is said to *reach v through U* iff there is a neighborhood N of $d(v)$ such that $N \cap d(\{v,w\}) \subseteq U \cup \{d(v)\}$.

3 A $\left(\left\lfloor \frac{3}{2}\mathrm{genus}(G)\right\rfloor + 3\right)$-Cop Capture Strategy

The idea for our strategy is quite simple: In [16] two cops are needed to reduce the genus of the robber territory by 1. This leads to the bound $2g + 3$ for the copnumber. Using fewer cops to reduce the genus of the robber territory must lead to better bounds. Our strategy will in each step produce either a strategy for one cop to reduce the genus of the robber territory by 1 or a strategy for two or three cops to reduce the genus of the robber territory by 2.

Our capture strategy has two natural parts: There is a recursive procedure that successively shrinks the robber territory until the genus drops by 1 or until only one cop is needed to keep the robber in the robber territory, given in Proposition 3.1. The heart of this procedure is a "leapfrogging strategy" executed by two cops to "flush" the robber "from left to right". The initialization (which already reduces the genus of the robber territory by 1) is given in the proof of Proposition 3.2. A visualization of its proof is given in Figure 3.1. Proposition 3.2 then easily leads to Theorem 3.3.

In both cases the visualization is quite natural, while the details become a little tedious. Looking at [1] seems to indicate that this is inherent in the subject, while Proposition 4.1 seems to indicate that the use of retractions presented here cannot be simplified easily.

Proposition 3.1 *Let $G = (V, E)$ be a finite connected graph of genus g with a good drawing d on the g-unit-handled plane such that*

1. *There are isometric paths P_1 and P_2 in G such that $d[P_1] = \{(x, 1) : -1 \le x \le 1\}$ ($d[P_1] = \{(1, 1)\}$ if $|P_1| = 1$) and $d[P_2] = \{(x, -1) : -1 \le x \le 1\}$ ($d[P_2] = \{(1, -1)\}$ if $|P_2| = 1$),*

2. *There are four cops γ_1, γ_2, γ_3 and γ_4 and a robber ρ on G,*

3. *For $i = 1, 2$ the cop γ_i has been moved to block P_i.*

If $g \ge 1$, there is a strategy such that ρ never enters any of P_1 or P_2 and such that eventually three cops or one cop restrict the movement of the robber to a subset T of the complement of $P_1 \cup P_2$. If three cops are needed, the genus of $G[T]$ is $\le g - 1$. If $g = 0$ there is a capture strategy on G using the given four cops such that ρ never enters any of P_1 or P_2.

Proof. The proof is an induction on the size of the robber territory $R = V \setminus (P_1 \cup P_2)$ (we can assume without loss of generality that the robber territory is the whole complement of $P_1 \cup P_2$). This is the set of vertices that the robber can reach without being captured on P_1 or P_2. The case $|R| = 0$ is trivial (in this case two cops suffice to capture). Note that since G is connected, some vertex in $P_1 \cup P_2$ is adjacent to a vertex in R.

Now assume that G is a finite connected graph of genus g with a good drawing d on a g-unit-handled plane S such that there are isometric paths $P_1 = \{p_0^1, \dots, p_{n_1}^1\}$ and $P_2 = \{p_0^2, \dots, p_{n_2}^2\}$ in G such that $d[P_1] = \{(x, 1) : -1 \le x \le 1\}$ ($d[P_1] = \{(1, 1)\}$ if $n_1 = 0$) and $d[P_2] = \{(x, -1) : -1 \le x \le 1\}$ ($d[P_2] = \{(1, -1)\}$ if $n_2 = 0$). Suppose there are four cops γ_1, γ_2, γ_3 and γ_4 and a robber ρ on G. Suppose for $i = 1, 2$ the cop γ_i has been moved to block P_i. Finally suppose that in all such configurations with robber territories with fewer vertices than $V \setminus (P_1 \cup P_2)$ the result holds.

Case 1: All vertices in $P_1 \cup P_2$ that are adjacent to vertices in R are on one path P_i. In this case, let γ_i continue blocking P_i. The cop that was blocking the other path is free to move. In case $g > 0$, we are done. In case

G is planar, use the three free cops (γ_3, γ_4 and the one that just became free to move) to capture the robber in the planar graph $G[R]$. This concludes the proof in Case 1.

From now on we can assume without loss of generality that no vertex in P_1 is adjacent to a vertex in P_2. Indeed, if this was the case, obtain G' from G by erasing all edges between vertices in P_1 and vertices in P_2 and find a strategy as desired in G'. Since the robber cannot reach any of the erased edges we are only creating additional constraints for the cop movement and any strategy in G' trivially induces a strategy in G. The case in which removal of such an edge disconnects G have been treated in Case 1.

Case 2: Vertices on both P_1 and P_2 are adjacent to vertices in R. If there is a vertex $v_1 \in P_1$ and a vertex $v_2 \in P_2$ with a geodesic path P_3 from v_1 to v_2 such that removal of $d[P_3]$ together with the vertical rays $\{d(v_1) + t(0,1) : t \geq 0\}$ and $\{d(v_2) + t(0,-1) : t \geq 0\}$ does not disconnect S, then the genus of $G[V \setminus (P_1 \cup P_2 \cup P_3)]$ is $\leq g - 1$. Moving γ_3 to block P_3 restricts the robber's movements to $G[T] := G[V \setminus (P_1 \cup P_2 \cup P_3)]$ and proves the result.

Otherwise let $p^1_{i_1}, \ldots, p^1_{i_k}$ be the vertices in P_1 (with $i_1 < i_2 < \ldots < i_k$) which are adjacent to a vertex in R and let $p^2_{j_1}, \ldots, p^2_{j_l}$ be the vertices in P_2 (with $j_1 < j_2 < \cdots < j_l$) which are adjacent to a vertex in R. Without loss of generality assume $k \leq l$.

For $z = 1, \ldots, k-1$ let Q_z be a geodesic path from $q^1_z := p^1_{i_z}$ to $q^2_z := p^2_{i_z}$ and for $z = k, \ldots, l$ let Q_z be a geodesic path from $q^1_z := p^1_{i_k}$ to $q^2_z := p^2_{i_z}$. By our earlier assumption, each Q_z has at least three vertices and thus at least one vertex in R. Let

$$L_z := d[Q_z] \cup \{d(q^1_z) + t(0,1) : t \geq 0\} \cup \{d(q^2_z) + t(0,-1) : t \geq 0\}.$$

By assumption, removal of any L_z disconnects S. The geodesic paths can be chosen in such a way that if $i < j$, then no points of $d[Q_j]$ are to the left of L_i. Move cop γ_3 to block Q_1.

If $d(\rho)$ is to the left of L_1, then γ_3 prevents ρ from reaching any vertex in P_1 or P_2. Thus γ_1 and γ_2 are free to move once more. If $g > 0$, this proves the result. If $g = 0$, the robber's movements are restricted to the planar graph drawn to the left of L_1. The other three free cops can capture the robber there.

In case $l = 1$ we have that $d(\rho)$ is either to the right or the left of L_1. Thus in this case the above argument (or its symmetric counterpart) finishes the proof.

In case $l > 1$, suppose $z < l$, a cop of γ_3 and γ_4, say γ_3 (which can always be achieved via renaming), has been moved to block Q_z and the movement of the robber is limited such that $d(\rho)$ must be to the right of L_z. This (with $z = 1$) is the only situation that remains to be treated in the previous discussion. Move γ_4 to block Q_{z+1}. If $d(\rho)$ is between L_z and L_{z+1}, then γ_1 and γ_2 are free to move once more. Let R' be the set of all vertices that the robber can reach in this configuration without being captured.

If Q_z and Q_{z+1} do not intersect, $G[R' \cup Q_z \cup Q_{z+1}]$ can be redrawn as is stated in the proposition and $|R'| < |R|$. Thus the result follows from the induction hypothesis.

If $Q_z =: \{a_1, \dots, a_n\}$ and $Q_{z+1} =: \{b_1, \dots, b_m\}$ intersect, obtain G' from G by doing the following. If $b_1 = a_1$, remove b_1 from Q_{z+1} and let γ_4 block the new Q_{z+1} only. Rename vertices and change m so that we again have $Q_{z+1} = \{b_1, \dots, b_m\}$. (If this is the only intersection we conclude as above.) For each maximal connected segment $b_{k_1}, \dots, b_{k_2}$ $(k_1 > 1)$ of Q_{z+1} that is contained in Q_z add a path $\tilde{b}_{k_1}, \dots, \tilde{b}_{k_2}$ and edges $\{b_{k_1-1}, \tilde{b}_{k_1}\}$ and $\{\tilde{b}_{k_1}, b_{k_1+1}\}$ to $G[R' \cup Q_z \cup Q_{z+1}]$. Let $\tilde{B}$ be the new path $x_1, \cdots, x_m$ obtained from $\{b_1, \dots, b_m\}$ by replacing every b_i for which there is a $\tilde{b}_i$ with $\tilde{b}_i$. $\tilde{B}$ is isometric in G'. G' can be drawn as stated in the proposition and the set of vertices the robber can reach without being captured is smaller than R. The map f that maps all $\tilde{b}_i$ to b_i and leaves all other vertices fixed is a map as desired in Lemma 2.4. A cop stationed at $f^{-1}(\gamma_4)$ blocks $\tilde{B}$, γ_3 still blocks Q_z. Thus the induction hypothesis can be applied to G'. Projecting the thus generated strategy back to G (cf. Lemma 2.4) proves the result.

If $d(\rho)$ is not between L_z and L_{z+1}, $d(\rho)$ is to the right of L_{z+1} and we can let γ_3 move freely once more. If $z + 1 < l$, continue the procedure described here. If $z + 1 = l$, then $d(\rho)$ is to the right of L_l. Continue symmetrically to the case when $d(\rho)$ was to the left of L_1. ■

Proposition 3.2 *Let G be a graph of genus $g \geq 1$ with four cops γ_1, γ_2, γ_3 and γ_4 and a robber ρ on it. Then two or three cops can be moved and assigned a strategy in such a way that the movement of the robber is restricted to a graph of genus $\leq g - 2$ or one cop can be moved and assigned a strategy such that the movement of the robber is restricted to a graph of genus $\leq g - 1$. If $g = 1$, there is a capture strategy on G with four cops.*

Proof. The proof will first show how to assign strategies to two cops to block certain cycles and second, how this configuration can be led into a recursion as described in Proposition 3.1.

Let d be a drawing of G on a surface S of genus g that contains the cylinder

$$\mathbb{Z} := \left\{ (x, y, z) : (x-2)^2 + z^2 = 1, -\frac{1}{2} \leq y \leq \frac{1}{2} \right\}.$$

Let $C = \{c_0, \dots, c_n\}$ be a cycle of minimum length such that $S \setminus d[C]$ still is connected (such a cycle exists, cf. [16], Lemma 1). Without loss of generality we can assume that $d[C] = \{(x, 0, z) : (x-2)^2 + z^2 = 1\}$ (every closed curve on S that does not intersect itself has a neighborhood that is isomorphic to an annulus), cf. Figure 3.1a). There are edges that reach

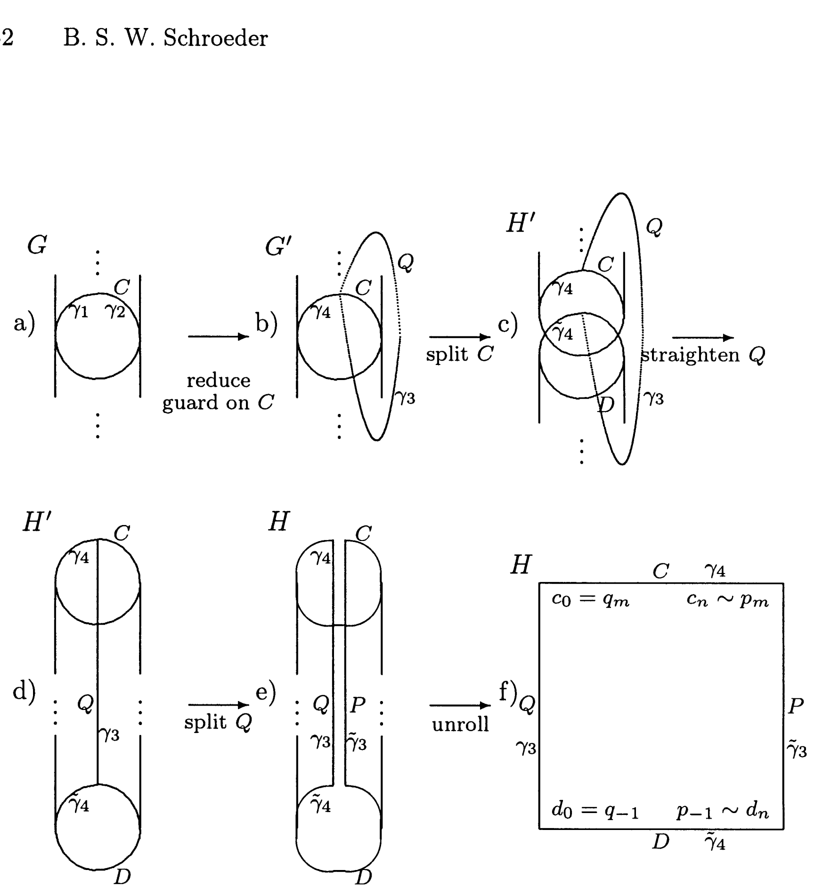


FIGURE 3.1. Illustration of the proof of Proposition 3.1. Each path or cycle is named as in the proof and has the names of the guarding cops next to it. Cops $\tilde{\gamma}_i$ are "virtual cops", i.e., inverse images of an actual cop under a quotient map.

C through the half-space $\{(x, y, z) : y > 0\}$, since otherwise removal of the circle $K := \{(x, \varepsilon, z) : (x-2)^2 + z^2 = 1\}$ for sufficiently small $\varepsilon > 0$ leaves us with a drawing of G on the surface $S \setminus K$, which has genus $g-1$, a contradiction. Similarly, there must be edges that reach C through the half-space $\{(x, y, z) : y < 0\}$. Let us assume without loss of generality that $d(c_0) = (2, 0, 1)$ and that there is an edge that reaches c_0 through the half-space $\{(x, y, z) : y < 0\}$.

C is the union of two geodesic paths $P_1 = \{c_0 = p_0^1, \dots, p_{n_1}^1\}$ and $P_2 = \{p_0^2, \dots, p_{n_2}^2\}$. For $i = 1, 2$, move γ_i to block P_i. Now the robber's moves are restricted to the graph $G[V \setminus C]$.

Let $q_0 \in V \setminus C$ be such that q_0 is adjacent to $p_0^1 = c_0$ and such that $\{q_0, p_0^1\}$ reaches p_0^1 through the half-space $\{(x, y, z) : y < 0\}$. Obtain G' from G by removing all edges that reach C through the half-space $\{(x, y, z) : y < 0\}$. Then the genus of G' is $\leq g - 1$. If genus$(G') \leq g - 2$, we are done since genus$(G[V \setminus C]) \leq$ genus(G'), so WLOG we can assume G' has genus

$g-1$. Let $Q = \{q_0, \dots, q_m = p_0^1 = c_0\}$ be a geodesic path in G', cf. Figure 3.1b). Move γ_3 to block Q in G'. This restricts the robber's moves to $G[V \setminus (C \cup Q)] = G'[V \setminus (C \cup Q)]$. Without loss of generality we can assume that

$$d[Q] \cap \mathbb{Z} = \left\{(2, y, 1) : -\frac{1}{2} \leq y \leq \frac{1}{2}\right\}.$$

Indeed, if $Q \cap C = \{c_0\}$ this is a simple re-drawing of Q (and a possible shifting of the drawings of some other points). If $|Q \cap C| \geq 2$, let k be the smallest index such that $q_k \in Q \cap C$. We can assume without loss of generality that all q_j with $j \geq k$ are in C and that $c_1 \notin Q$. Obtain $\tilde{G}$ from G' as follows (the same operations would also have to be performed on G):

1. Add a path $q'_k, \dots, q'_{m-1}$ and edges $\{q_{k-1}, q'_k\}$ and $\{q'_{m-1}, q_m\}$ to G'.
2. For $k \leq i \leq m-1$, remove all edges $\{v, q_i\}$ that reach q_i through the half-space $\{(x, y, z) : y > 0\}$ and replace them with edges $\{v, q'_i\}$.

Then $\tilde{G}[V \setminus (Q' \cup C)] = G'[V \setminus (Q \cup C)]$ and $\tilde{G}$ has a drawing $\tilde{d}$ on S such that the geodesic path $Q' = \{q_0, \dots, q_{k-1}, q'_k, \dots, q'_{m-1}, q_m\}$ satisfies

$$\tilde{d}[Q'] \cap \mathbb{Z} = \left\{(2, y, 1) : -\frac{1}{2} \leq y \leq \frac{1}{2}\right\}.$$

We could now continue with $\tilde{G}$ instead of G', eventually using Lemma 2.4 to project the designed strategies back to G'. Instead, we assume WLOG that G, G' and d have the desired property.

$S \setminus d[G[Q]]$ must be connected, since for every ε-neighborhood $N_\varepsilon(2,0,1)$ of $d(c_0) = (2,0,1)$, $d[C] \setminus N_{\frac{\varepsilon}{2}}(2,0,1)$ is a path that in the intersection of $N_\varepsilon(2,0,1)$ with $S \setminus d[G[Q]]$ connects points with $x < 2$ with points with $x > 2$, cf. Figure 3.1b).

Obtain G'' from G by removing the vertices $q_0, \cdots, q_{m-1}$ and all edges that reach $p_0^1 = q_m = c_0$ through the half-space $\{(x,y,z) : x > 2\}$. Without loss of generality we can assume that the edge $\{c_0, c_1\}$ is in G''. The path $G''[\{c_0, c_1, \dots, c_n\}]$ is geodesic in G''. Indeed, otherwise there is a $Z := \{z_0 = c_0, z_1, \dots, z_k = c_n\}$ with $k < n$ such that $G[Z]$ is a cycle and such that $d[G[Z]]$ intersects $d[G[Q]]$ in exactly one point, namely c_0. But then $S \setminus d[G[Z]]$ must be connected, since for every $\varepsilon > 0$, $d[G[Q]] \setminus N_{\frac{\varepsilon}{2}}(2,0,1)$ is a path that in the intersection of $N_\varepsilon(2,0,1)$ with $S \setminus d[G[Z]]$ connects points with positive y-coordinate with points with negative y-coordinate. This would be a contradiction to the minimality of C.

For $v \in (V \setminus Q) \cup \{c_0\}$ let

$$r''_{C,c_0}(v) := \begin{cases} c_{d_{G''}(v,c_0)}; & \text{if } d_{G''}(v,c_0) \leq n, \\ c_n; & \text{if } d_{G''}(v,c_0) > n. \end{cases}$$

Move γ_4 to $r''_{C,c_0}(\rho)$. Since the movement of ρ was restricted to $G[V \setminus (C \cup Q)] = G''[V \setminus (C \cup Q)]$ the cop γ_4 suffices now to keep ρ from entering C and γ_1 and γ_2 are free to move away from C once more, cf. Figure 3.1b).

This ends the assignment of strategies to the cops. In the following we will show how to use the methods of Proposition 3.1 to finish the proof. We could have translated the methods of Proposition 3.1 directly to this situation. However this approach would have led to technical difficulties when defining what "between", "to the left" and "to the right" should mean on S. Moreover we still would have needed Proposition 3.1 for the arising recursion. Thus the author decided to present the "unrolling of G" described in the following.

Define the graph H' by adding to G' vertices $d_0, \dots, d_n$, edges $\{d_i, d_{i+1}\}$, $i = 0, \dots, n-1$, and edges $\{v, d_i\}$ for all v such that in G, the edge $\{v, c_i\}$ reaches c_i through the half-space $\{(x,y,z) : y < 0\}$. Obtain a drawing of H' on S by placing the drawings of the d_i at $d(c_i) - (0, \varepsilon, 0)$ for sufficiently small $\varepsilon > 0$ and by putting in the drawings of the edges in the obvious way. Call this drawing d', cf. Figure 3.1c,d).

Let $q_{-1} := d_0$. Choose $\varepsilon > 0$ small enough so that no drawing of any vertex of $V_{H'} \setminus (Q \cup \{q_{-1}\})$ is in the ε-neighborhood of $d'[H'[Q \cup \{q_{-1}\}]]$ and such that all the edges that enter the ε-neighborhood terminate at a vertex q_i. Draw a circle K' in the union of $N_\varepsilon(d(\{c_0, q_0\}))$ and the ε-neighborhood of $d'[H'[Q \cup \{q_{-1}\}]]$ such that

1. K' does not intersect $d'[H'[Q \cup \{q_{-1}\}]]$,
2. For every point x of $d'[H'[Q \cup \{q_{-1}\}]]$, there is a point of K' that is within ε of x,
3. K' intersects the drawing of every edge at most once,
4. K' intersects the ε-neighborhood of $(2,0,1)$ at some point with $x > 2$.

(I.e., K' is a circle "just to the right" of $d'[H'[Q \cup \{q_{-1}\}]]$.) Define a graph H by erasing from H' all edges that intersect the circle K', adding vertices $p_{-1}, \dots, p_m$, edges $\{p_i, p_{i+1}\}$, $i = -1, \dots, m-1$, and edges $\{v, p_i\}$, $i = -1, \dots, m$, for all v such that the drawing $d'(\{v, q_i\})$ of the edge $\{v, q_i\}$ intersects K'. Also add the edges $\{d_n, p_{-1}\}$ and $\{c_n, p_m\}$. Obtain a drawing d_H of H on S from d' in the following way: Erase all the edges that were removed, draw every p_i at the point on K' that is closest to q_i (if K' is chosen properly such a point is unique and two such points for distinct i are distinct) and then draw the new edges in the obvious way, cf. Figure 3.1e).

Let $D := \{d_0, \cdots, d_n\}$ and $P := \{p_{-1}, p_0, \dots, p_m\}$. Then $d_H[C \cup P \cup D \cup Q]$ is isomorphic to a circle and removal of $d_H[C \cup P \cup D \cup Q]$ disconnects S into a piece in which no part of H is drawn and another piece of genus $\tilde{g} \le (g-1)$ (recall the genus of G' is $g-1$). For a visualization cf. Figure 3.1b)–f). Algebraically we would need to parametrize the paths $d(D)$, $d(P)$ and $d(Q)$ and use that WLOG they could all be drawn on the union of $\mathbb{Z}$ with $\left\{(x,y,z) : y^2 + (z+1)^2 = 4, -\frac{1}{2} \leq x \leq \frac{1}{2}\right\}$. (With small neighborhoods of $(2,0,1)$ on either cylinder identified.)

If $\tilde{g} \leq (g-2)$, we have placed two cops such that the movement of the robber is restricted to a graph of genus $\leq (g-2)$ and we are done.

If $\tilde{g} = (g-1)$, there is a drawing d_H^u of H on a $(g-1)$-unit-handled plane such that no part of H is drawn outside the unit square and such that $d_H^u[C] = \{(x,y) : -1 \leq x \leq 1, y = 1\}$, $d_H^u[D] = \{(x,y) : -1 \leq x \leq 1, y = -1\}$, $d_H^u[Q \cup \{q_{-1}\}] = \{(x,y) : x = -1, -1 \leq y \leq 1\}$ and $d_H^u[P] = \{(x,y) : x = 1, -1 \leq y \leq 1\}$, cf. Figure 3.1f).

Consider the quotient map from H to G defined by

$$F(v) := \begin{cases} v; & \text{if } v \notin D \cup P, \\ c_i; & \text{if } v = d_i \in D, \\ q_i; & \text{if } v = p_i \in P, i \neq= -1, \\ c_0; & \text{if } v = p_{-1}. \end{cases}$$

Unless the robber moves onto C or Q (in which case the robber is immediately captured by γ_4 or γ_3), the robber ρ will have a unique preimage $F^{-1}(\rho)$. Thus a stationary cop at p_{-1} (needed only when γ_3 is at c_0, which is the only time γ_3 has two preimages in P) and two preimages of γ_3, one each on P and $Q \cup \{d_0\}$, prevent $F^{-1}(\rho)$ from entering P and $Q \cup \{d_0\}$. The two preimages of γ_4, one each on C and D, prevent $F^{-1}(\rho)$ from entering C and D. (We could say that two cops and three "virtual cops" block $F^{-1}(\rho)$ from reaching $C \cup P \cup D \cup Q$, i.e., $a = 5$ in Lemma 2.4.)

We are thus in a situation in which on H we can use cops γ_1 and γ_2 to run the strategy outlined in Case 2 of the proof of Proposition 3.1. Let C play the role of P_1 and let D play the role of P_2. We may obtain a path from a vertex of C to a vertex of D such that removal of this path decreases the genus of H by 1. In this case we have assigned three cops to limit the robber's moves to a graph of genus $\leq g-2$ and can project the strategy back to G via Lemma 2.4. Otherwise we will (by starting with γ_1 also blocking $Q \cup \{q_{-1}\}$ and leapfrogging as described in the proof of Proposition 3.1, ending if necessary with one of γ_1, γ_2 blocking P) generate a situation as described in Proposition 3.1 with the robber trapped between two paths guarded by γ_1 and γ_2 and cops γ_3 and γ_4 free to move once more. Project the thus obtained placement or capture strategy back to G via Lemma 2.4. Now we are able to apply Proposition 3.1 directly. ∎

Theorem 3.3 *Let G be a finite connected graph of genus g. Then the copnumber of G is bounded by* $\left\lfloor \frac{3}{2} g \right\rfloor + 3$.

Proof: For $g = 0$ this is Aigner and Fromme's result (cf. Theorem 6 in [1]). For $g = 1$, by Proposition 3.2 four cops suffice. Now suppose $g >$

1. Repeatedly apply Proposition 3.2 until the movement of the robber is restricted to a graph of genus 1 or less, then use a capture strategy with three or four cops (depending on whether the robber is eventually restricted to moving on a planar or a toroidal graph).

Let k be the number of times three cops were used to reduce the genus of the robber territory by at least 2 and let l be the number of times one cop was used to reduce the genus of the robber territory by at least 1. To simplify notation we count a use of two cops to reduce the genus of the robber territory by at least 2 as two uses of one cop to decrease the genus of the robber territory by at least 1 each. In case $2k + l = g$, we have $3k + \frac{3}{2}l = \frac{3}{2}g$. The number of cops used in the capture strategy is $3k + l + 3 \leq \left\lfloor 3k + \frac{3}{2}l \right\rfloor + 3 = \left\lfloor \frac{3}{2}g \right\rfloor + 3$. In case $2k + l \leq g - 1$, we have $3k + \frac{3}{2}l \leq \frac{3}{2}g - \frac{3}{2}$. The number of cops used in the capture strategy is at most $3k + l + 4 \leq \left\lfloor 3k + \frac{3}{2}l \right\rfloor + 4 \leq \left\lfloor \frac{3}{2}g \right\rfloor - 1 + 4 = \left\lfloor \frac{3}{2}g \right\rfloor + 3$. ■

4 Remarks on a Possible Bound of Genus+3

While bounding the copnumber of a graph of genus g by $\left\lfloor \frac{3}{2}g \right\rfloor + 3$ is an improvement over the original bound of $2g+3$, the author conjectures that the copnumber of a graph of genus g is at least bounded by $g+3$. In Section 5 we support this conjecture by sketching a proof that the copnumber of a graph of genus 2 is bounded by 5. In [16] it is necessary to use two cops each to block the robber's access to a non-null-homotopic cycle. Since up to g non-null-homotopic cycles need to be blocked and three cops are needed to effect capture in the remaining planar graph, this leads to the estimate of $2g + 3$. The obvious idea to prove the copnumber is in fact bounded by $g + 3$ is to attempt to use one cop to block a non-null-homotopic cycle. With the tools presently available this is only possible if non-null-homotopic cycles are retracts. Proposition 4.1 shows that non-null-homotopic cycles need not be retracts, thus destroying the hope for a quick resolution of the conjecture.

Proposition 4.1 *There is a toroidal graph $G = (V, E)$ such that no non-null-homotopic (with respect to any drawing of G) cycle of G is a retract.*

Proof. Let $G = (V, E)$ with $V = \{1, \ldots, n\} \times \{1, \ldots, n\}$, $n \geq 20$, n even and

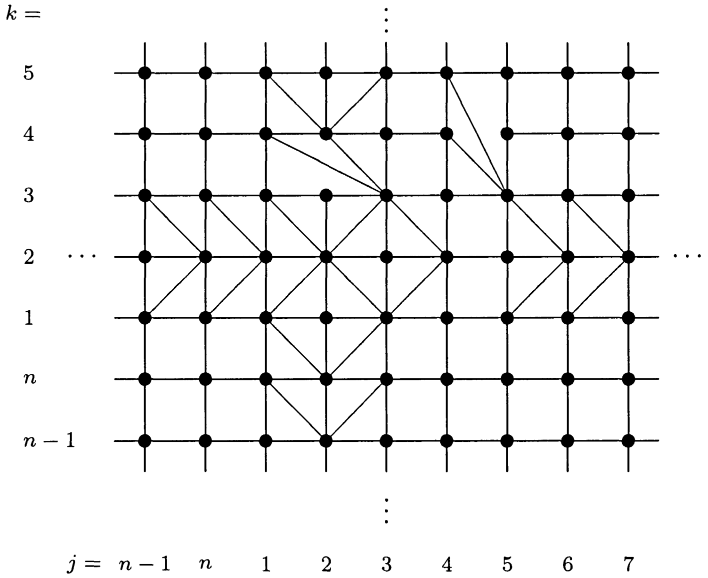


FIGURE 4.2. The graphs in Proposition 4.1.

$$
\begin{aligned}
E \;=\; & \{\{(j,k),(j+1,k)\} : j,k = 1,\dots,n \bmod n\} \setminus \{\{(4,4),(5,4)\}\} \cup \\
& \{\{(j,k),(j,k+1)\} : j,k = 1,\dots,n \bmod n\} \setminus \{\{(2,3),(2,4)\}\} \cup \\
& \{\{(2,k),(3,k+1)\},\{(2,k),(1,k+1)\} : k = 4,\dots,n \bmod n\} \cup \\
& \{\{(j,3),(j+1,2)\},\{(j,1),(j+1,2)\} : j = 5,\dots,n,1 \bmod n\} \cup \\
& \{\{(3,1),(2,2)\},\{(2,2),(3,3)\},\{(3,3),(1,4)\}\} \cup \\
& \{\{(3,3),(2,4)\},\{(5,3),(4,4)\},\{(5,3),(4,5)\}\} \cup \\
& \{\{(4,2),(3,3)\},\{(4,2),(3,1)\}\};
\end{aligned}
$$

cf. Figure 4.2 for an illustration.

Clearly G is toroidal, as G is essentially the cartesian product of two n-cycles with small modifications only. To begin with, as drawing d we choose the drawing on a torus as indicated in Figure 4.2. All references to homotopy will be with respect to the drawing d. We will first show that no cycle that is not null-homotopic with respect to this drawing of G is a retract, generalizing to arbitrary drawings later.

First note that all cycles $\{\{j,k\} : j = 1,\dots,n\}$, $k \in \{1,\dots,n\} \setminus \{4\}$ and $\{\{j,k\} : k = 1,\dots,n\}$, $j \in \{1,\dots,n\} \setminus \{2\}$ are minimal non-null-homotopic cycles of G.

Step 1: Cycles in which one coordinate is fixed cannot be retracts. Suppose there was a retraction $r : V \to \{\{j,1\} : j = 1,\dots,n\}$. Then $r((j,2)) \in \{(j-1,1),(j,1),(j+1,1)\} \pmod n$ since $(j,2) \sim (j,1)$. This implies $r|_{\{(j,2):j=1,\dots,n\}}$ is injective. Since $(2,2) \sim (3,1)$, we have

$r((2,2)) \neq (1,1)$. Thus $r((j,2)) \neq (j-1,1) \pmod n$. Since $(4,2) \sim (3,1)$, we have $r((4,2)) \neq (5,1)$. Thus $r((j,2)) \neq (j+1,1) \pmod n$. Hence (*) $r((j,2)) = (j,1)$ for all $j = 1, \ldots, n$.

Similarly we prove $r((j,3)) = r((j,2)) = (j,1)$ ($r((3,3))$ must be adjacent or equal to $(2,1), (3,1)$ and $(4,1)$, since $(3,3)$ is adjacent to $(2,2), (3,2)$ and $(4,2)$, thus $r((3,3)) = (3,1)$).

Now $r((j,4)) \in \{(j-1,1), (j,1), (j+1,1)\}$ for $j \neq 2 \pmod n$ since $(j,4) \sim (j,3)$ for $j \neq 2$. This implies $r|_{\{(j,4):j=1,\ldots,n\}}$ is injective and $r((2,4)) \in \{(1,1),(2,1),(3,1)\}$. Since $(1,4) \sim (3,3)$, we have $r((1,4)) \notin \{(1,1),(n,1)\}$. Thus $r((j,4)) \notin \{(j-1,1),(j,1)\} \pmod n$. Hence $r((j,4) = (j+1,1)$ for all $j = 1, \ldots, n$.

Similar to (*) we can now prove that $r((j,k)) = r((j,k-1)) = (j+1,1)$ for $k = 5, \ldots, n$. However then $(2,n) \sim (1,1)$ and $r((2,n)) = (3,1) \not\sim (1,1) = r((1,1))$, a contradiction.

Thus G has no retraction onto $\{\{j,1\} : j = 1, \ldots, n\}$. Similarly we prove G has no retraction onto any $\{\{j,k\} : j = 1, \ldots, n\}$, $k \in \{2, \ldots, n\} \setminus \{4\}$ and that G has no retraction onto any $\{\{j,k\} : k = 1, \ldots, n\}$, $j \in \{1, \ldots, n\} \setminus \{2\}$.

Step 2: No minimal non-null-homotopic cycle is a retract. Now let π_1 and π_2 be the projections on the first and second coordinates respectively. If C is another minimal non-null-homotopic cycle, one checks that either there is a j_0 such that $\pi_1[V(C)] \in \{j_0 - 1, j_0, j_0 + 1\} \pmod n$ or there is a k_0 such that $\pi_2[V(C)] \in \{k_0 - 1, k_0, k_0 + 1\} \pmod n$. If G had a retraction s onto $G[C]$, then each $c \in C$ would have a unique preimage in $\{(j,k_0) : j = 1, \ldots, n\}$ or a unique preimage in $\{(j_0,k) : k = 1, \ldots, n\}$. (If $k_0 = 4$ or $j_0 = 2$, replace k_0 with 3 or j_0 with 1 and the statement still holds.) Composing s with the thus induced isomorphism between $G[C]$ and $G[\{(j,k_0) : j = 1, \ldots, n\}]$ or $G[\{(j_0,k) : k = 1, \ldots, n\}]$ produces a retraction onto $G[\{(j,k_0) : j = 1, \ldots, n\}]$ or $G[\{(j_0,k) : k = 1, \ldots, n\}]$, a contradiction.

Step 3: No cycle of length $> n+1$ is a retract. Suppose $G[K] = G[\{x_1, \ldots, x_m\}]$ is a cycle with $m > n+1$ elements and t is a retraction of G onto the cycle $G[K]$. Then for each cycle $A_k := \{(j,k) : j = 1, \ldots, n\}$, $k \in \{1, \ldots, n\}$ (for $k = 4$ we add the vertex $(5,3)$ to obtain a cycle) and $B_j := \{(j,k) : k = 1, \ldots, n\}$, $j \in \{1, \ldots, n\}$ (for $j = 2$ we add the vertex $(3,3)$ to obtain a cycle) we have that (since n was even) $t[A_k]$ and $t[B_j]$ is a path with at most $\frac{n}{2} + 1$ vertices.

For $i, j \in \{1, \ldots, m\}$ let $K(i,j)$ be the path $x_i, x_{i+1}, \ldots, x_{j-1}, x_j$ if $i \leq j$ and the path $x_i, x_{i+1}, \ldots, x_m, x_1, \ldots, x_{j-1}, x_j$ if $i > j$. Then for all $k \in \{1, \ldots, n\}$ we have that $t[A_k] = K(l_k, r_k)$ for some $l_k, r_k \in \{1, \ldots, m\}$. Since $m > n+1$, there is an $i \in \{1, \ldots, m\}$ such that for all $k \in \{1, \ldots, n\}$ we have that $x_i \neq l_k$. Without loss of generality we can thus assume that $x_m \neq l_k$ for all k and that $x_1 = l_{k_0}$ for some $k_0 \in \{1, \ldots, n\}$.

Let j_0 be such that $t((j_0, k_0)) = x_1$. (In case $k_0 = 4$ and no such j_0 exists, we must have $t((5,3)) = x_1$ and we redefine $k_0 = 3$ and let $j_0 = 5$.)

The path $t[B_{j_0}]$ has length $\leq \frac{n}{2}$. Suppose $x_m \in t[B_{j_0}]$. Let k_1 be the first number k in the sequence $k_0, k_0+1, \dots, n, 1, \dots, k_0 - 1$ such that $t((j_0, k)) = x_m$. Then $l_{k_1} \in \left\{m - \frac{n}{2}, \dots, m\right\}$ and for all $k \in \{k_0, \dots, k_1\}$ we have $t((j_0, k)) = x_i$ for some $i \in \left\{l_k, \dots, l_k + \frac{n}{2}\right\}$. Since B_{j_0} is connected this would mean that $t[B_{j_0}]$ has m elements, a contradiction.

Thus $t[B_{j_0}] \subseteq K\left(1, \frac{n}{2} + 1\right)$. Let $i(j_0, k)$ be the index such that $t((j_0, k)) = x_{i(j_0,k)}$. Then $t[A_k] \subseteq K\left(1, i(j_0, k) + \frac{n}{2}\right)$ (and in fact $i(j_0, k) + \frac{n}{2}$ is only in $t[A_k]$ if $t[A_k]$ has $\frac{n}{2} + 1$ elements). However then $t[V] \subseteq K(1, n+1)$ and K has $\leq n + 1$ vertices, contradiction.

Moreover, no non-null-homotopic path of length $n + 1$ can be a retract, since this would (with an argument similar to step 2) allow us to construct a retraction onto a cycle in which one coordinate is fixed.

Thus G has no retraction onto any cycle of length $> n + 1$ and no retraction onto any non-null-homotopic cycle of length n or $n + 1$.

Step 4: Independence of the result from the chosen drawing. Now consider an arbitrary drawing d_a of G. Let T be the torus obtained by rotating the circle C_0 of radius 1, centered at $(2, 0, 0)$ about the z-axis. Without loss of generality we can assume that the range of d_a is $T \cup A(T)$. A_1 cannot be null-homotopic in any drawing of G, so we can assume without loss of generality that the range of $d_a[A_1]$ is $C_0 \cup A(C_0)$. B_1 cannot be null-homotopic in any drawing of G either and must (due to the position of $d_a[A_1]$) wrap around the origin. However this means that d_a and d have the same null-homotopic cycles, which finishes the proof. ■

In fact, the author conjectures that the graphs in the proof of Proposition 4.1 have no nontrivial nonplanar retract, which then in turn would imply that no nontrivial retract of the graphs in the proof of Proposition 4.1 contains a non-null-homotopic cycle.

Another idea to prove the copnumber is bounded by $g + 3$ would be to extend the strategy presented here, by blocking all paths Q_z in the proof of Proposition 3.1 such that removal of L_z does not disconnect the surface S and hope that this would further decrease the genus of the graph the robber can reach in such a fashion that each additional cop deployed decreases the genus by at least 1. Figure 4.3 shows that this need not be the case. We use the notation of the proof of Proposition 3.1. One representative handle is drawn into the picture, but other handles that connect the territory on the left with the territory on the right might exist. The paths Q_i and Q_j guarded by cops γ_3 and γ_4 are both such that removal of L_i or L_j does not disconnect the surface S. Removal of $L_i \cup L_j$ does disconnect S and the genus of the component that contains the robber can still be $g - 1$ (if

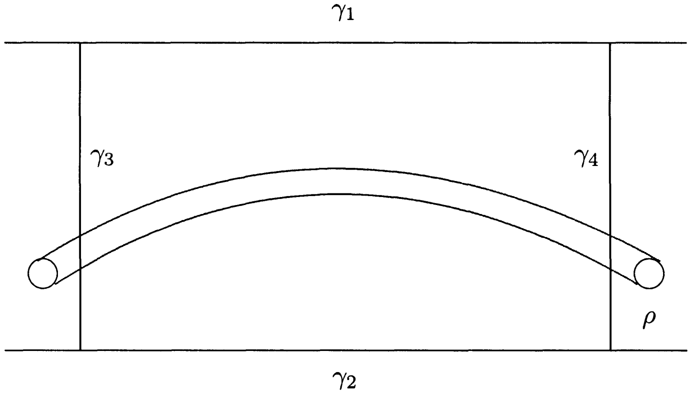


FIGURE 4.3. A situation in which a natural extension of the strategy in Proposition 3.1 which gives the $\left\lfloor \frac{3}{2}g \right\rfloor + 3$ bound fails.

the other component is indeed planar). None of the cops is free to move as they all guard pieces of paths that the robber could reach, were they unguarded. Finally it is not possible to use γ_3 and γ_4 as the outer paths in a continuation, as it need not be possible to re-draw the robber territory plus bounding paths in such a fashion that γ_3 guards a path drawn at the top side of the unit square and γ_4 guards a path drawn at the bottom side of the unit square.

5 The Copnumber of Graphs of Genus 2 is Bounded by 5

In this section we briefly indicate how one can prove that a graph of genus 2 has at most copnumber 5.

Theorem 5.1 *Let G be a graph of genus* 2. *Then the copnumber of G is at most* 5.

Proof (sketch). Start with the graph H as in the proof of Proposition 3.2 and with cops γ_3 and γ_4 (with their corresponding preimages on the other path) blocking $C \cup P \cup D \cup Q$. If the robber territory is already planar, use the three remaining cops to capture the robber. Otherwise the situation is as depicted at the top of Figure 5.4. Use γ_5 to block P and start leapfrogging towards the middle as in Proposition 3.1. When a path needs to be blocked such that removal of the corresponding L_k does not disconnect the surface, leave the cops blocking the last two paths assigned to those paths and start leapfrogging in from the other side. This will eventually lead to the robber being trapped in a planar territory between two guarded paths (in which

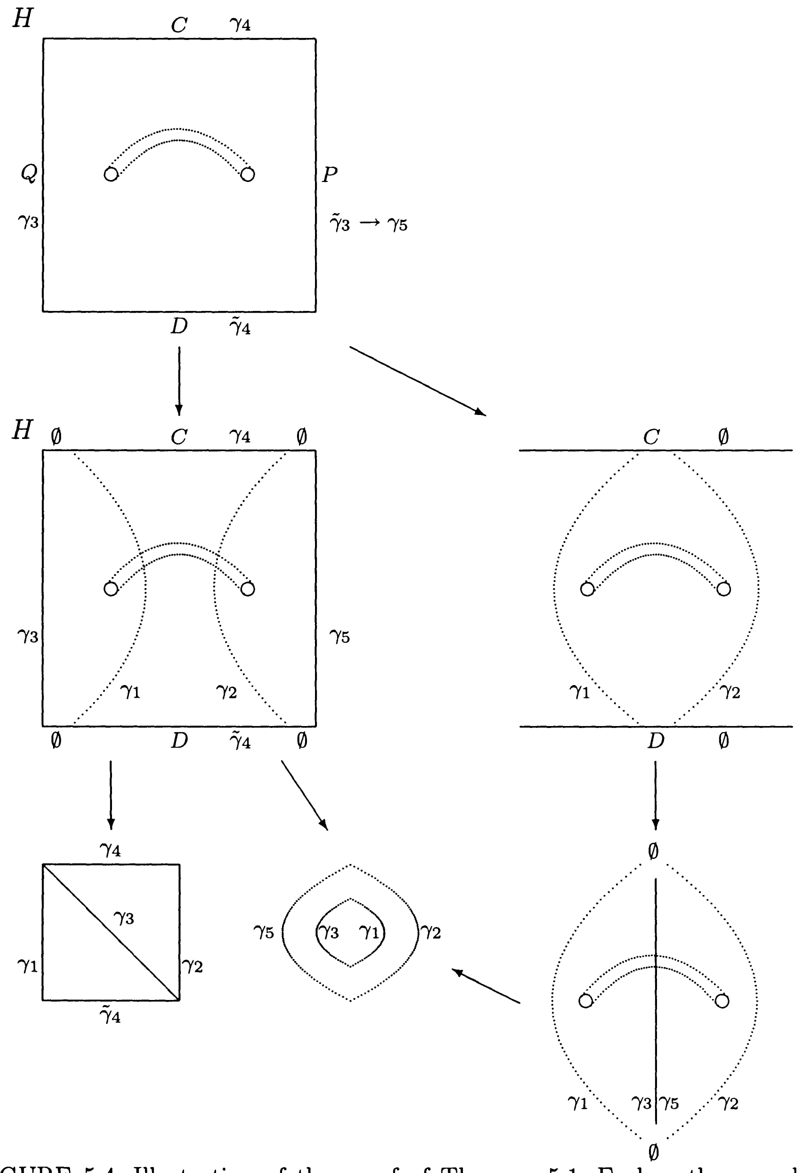


FIGURE 5.4. Illustration of the proof of Theorem 5.1. Each path or cycle is named as in the proof and has the names of the guarding cops next to it. Cops $\tilde{\gamma}_i$ are "virtual cops", i.e., inverse images of an actual cop under a quotient map. An $\emptyset$ next to a path or a segment of a path indicates that the path or segment no longer needs to be guarded.

case we are done), or to a configuration as shown in the two middle images in Figure 5.4 or to a configuration as shown on the bottom right (with different cop assignments than shown).

In the left middle image the robber could be in two possible territories. First, the territory whose boundaries are guarded by $\gamma_1, \gamma_2, \gamma_4$ and $\tilde{\gamma}_4$, leading to the bottom left situation. Second, the territory whose boundaries are guarded by $\gamma_1, \gamma_2, \gamma_3$ and γ_5, leading to the bottom middle situation.

In the right middle image we continue arguing similar to Proposition 3.1 until the removal of the drawing of a path does not disconnect the surface. This leads (after possibly renaming some cops and shortening the middle path) to the situation in the bottom right image in Figure 5.4. Doubling the guard on the middle path and making the obvious cut leads to the situation in the bottom middle image.

In the bottom middle image we can now keep γ_2 and γ_5 assigned to their tasks and start a planar capture strategy as outlined in [1] with γ_1 and γ_3 blocking the initial cycle (γ_4 is the third cop for this strategy). In the bottom left image we first block a geodesic that connects opposing corners. Now the robber is in one of the triangles. Use one of the blocked paths as one half of a cycle and block a geodesic from endpoint to endpoint in the robber territory. This leads to a blocked cycle that can be used to start a capture strategy as given in [1]. ■

Acknowledgement

I would like to thank R. Nowakowski for valuable discussions on this subject and his careful comments during a presentation of earlier results.

References

[1] M. Aigner and M. Fromme. A game of cops and robbers, *Discrete Appl. Math.*, **8** (1984), 1–12.

[2] T. Andreae. Note on a pursuit game played on graphs, *Discrete Appl. Math.*, **9** (1984), 111–115.

[3] T. Andreae. On a pursuit game played on a graph for which a minor is excluded, *J. Combin. Theory (Ser. B)*, **41** (1986), 37–47.

[4] R. Anstee and M. Farber. On bridged graphs and cop-win graphs, *J. Combin. Theory (Ser. B)*, **44** (1988), 22–28.

[5] A. Berarducci and B. Intrigila. On the copnumber of a graph, *Advances in Applied Mathematics*, **14** (1993), 389–403.

[6] S. Fitzpatrick and R. Nowakowski. The copnumber of graphs with strong isometric dimension 2, preprint (1997).

[7] P. Frankl. Cops and robbers in graphs with large girth and Cayley graphs, *Discrete Appl. Math.*, **17** (1987), 301–305.

[8] A. Goldstein and E. Reingold. The complexity of pursuit on a graph, *Theoretical Computer Science*, **143** (1995), 93–112.

[9] M. W. Hirsch. Differential Topology, *Graduate Texts in Mathematics*, **33**, Springer Verlag, New York, 1976.

[10] M. Maamoun and H. Meyniel. On a game of policemen and robber, *Discrete Appl. Math.*, **17** (1987), 307–309.

[11] S. Neufeld and R. Nowakowski. A vertex-to-vertex pursuit game with disjoint sets of edges, in: Sauer, et. al., *Finite and Infinite Combinatorics in Sets and Logic*, NATO ASI Series C, **Vol. 411** (1991), 299–312.

[12] R. Nowakowski, B. Schröder. Bounding the copnumber using the crosscap number, unpublished manuscript, 1998.

[13] R. Nowakowski and P. Winkler. Vertex-to-vertex pursuit in a graph, *Discrete Mathematics*, **43** (1983), 235–239.

[14] N. Polat, Retract-collapsible graphs and invariant subgraph properties, *J. Graph Theory*, **19**, (1995), 25–44.

[15] A. Quilliot. *Thèse de doctorat d'etat*, Université de Paris VI, 1983.

[16] A. Quilliot. *A short note about pursuit games played on a graph with a given genus*, J. Combin. Theory (Ser. B), **38** (1985), 89–92.

[17] A. White and L. Beineke. *Topological graph theory, in: Beineke and Wilson, Selected Topics in Graph Theory*, Academic Press, London, 1978.

Bernd S. W. Schröder
Program of Mathematics and Statistics
Louisiana Tech University
Ruston, LA 71272 USA
schroder@engr.LaTech.edu

Abelian Groups: Simultaneously Reflective and Coreflective Subcategories versus Modules

Robert El Bashir,[1] Horst Herrlich and Miroslav Hušek[2]

To our friends, Lamar Bentley and George Strecker, on the occasion of their 60th birthday – with affection and gratitude

ABSTRACT We investigate full subcategories of the category Ab of Abelian groups that are simultaneously reflective and coreflective in Ab. Such subcategories are exactly those isomorphic to categories of modules that are fully embedded into Ab. Rings giving rise to such modules are completely described. One of the curious special cases is provided by the full subcategory of Ab consisting of all torsion–free, divisible Abelian groups, which can be characterized, alternatively as the reflective hull of $\mathbb{Q}$ in Ab, or as the coreflective hull of $\mathbb{Q}$ in Ab, or as the intersection of the epireflective hull of $\mathbb{Q}$ in Ab with the monocoreflective hull of $\mathbb{Q}$ in Ab.

1 Introduction

Full subcategories $\mathcal{A}$ of $\mathcal{C}$ that are simultaneously reflective and coreflective in $\mathcal{C}$ inherit many desirable properties from $\mathcal{C}$, such as completeness, co-completeness, wellpoweredness, co-wellpoweredness, cartesian-closedness, Abelianness, etc. Thus, for a given category $\mathcal{C}$ with pleasant features, it is natural to ask for a description of all simultaneously reflective and coreflective full subcategories of $\mathcal{C}$. Such investigations have been carried out, e.g., for the categories Top of topological spaces [8], Unif of uniform spaces [6], [7], Bor of bornological spaces [3], AP of approach spaces [5], and various relational [4] and algebraic constructs. However, for the beautiful category

[1]The author acknowledges the support of the Czech Republic grant 201/98/P247 and MSM 113200007.

[2]The author acknowledges the support of the Czech Republic grant 201/00/1466 and MSM 113200007.

$\mathtt{Ab}$ of Abelian groups very little seems to be known so far. This paper is designed to fill this gap.

We use the terminology from [1] for categorical concepts and from [2] for group-theoretic concepts. Moreover, in this paper all groups are assumed to be Abelian and all subcategories to be full and closed under isomorphisms. If $\mathcal{C}$ is a reflective or coreflective subcategory of $\mathtt{Ab}$, we denote by r or c the respective reflector or coreflector $\mathtt{Ab} \to \mathcal{C}$ and $\{r_G\}$ or $\{c_G\}$ the respective natural transformations $1 \to r$ or $c \to 1$. If $f : G \to H$ in $\mathtt{Ab}$ for $H \in \mathcal{C}$, we shall denote by f_r the unique homomorphism $r(G) \to H$ from $\mathcal{C}$ with $f_r \circ r_G = f$, and call it *extension* of f. By tG we mean the torsion part of a group G and by $\mathbb{N}$, the set of positive integers will be denoted.

2 General Result

Theorem 2.1 *A subcategory $\mathcal{C}$ of $\mathtt{Ab}$ is simultaneously reflective and coreflective in $\mathtt{Ab}$ if and only if there exists a commutative unitary ring R such that the forgetful functor $U : \mathtt{Mod}_R \to \mathtt{Ab}$ is a full embedding onto $\mathcal{C}$.*

Proof. Suppose first that we have a commutative unitary ring R and $U : \mathtt{Mod}_R \to \mathtt{Ab}$ is the forgetful functor.

Both $\mathtt{Mod}_R$ and $\mathtt{Ab}$ are algebraic constructs via their canonical forgetful functors into $\mathtt{Set}$. By [1, Proposition 23.22], the functor U, being a concrete functor between algebraic constructs, is algebraic. Hence by [1, Proposition 23.8], U is a (right) adjoint.

$\mathtt{Mod}_R$ is cocomplete and cowellpowered, has R as a separator and, moreover, is closed under the formation of all colimits in $\mathtt{Ab}$. Hence by the dual of the Special Adjoint Functor Theorem (see, e.g., [1, Proposition 18.17]), the embedding $\mathtt{Mod}_R$ into $\mathtt{Ab}$ is coadjoint (i.e., a left adjoint).

Thus, if U is a full embedding, its image (i.e., $\mathcal{C}$ in our case) is both reflective and coreflective in $\mathtt{Ab}$.

Now, let $\mathcal{C}$ be a subcategory of $\mathtt{Ab}$ that is simultaneously reflective and coreflective in $\mathtt{Ab}$.

Claim 1. *The reflection $R = r\mathbb{Z}$ of $\mathbb{Z}$ in $\mathcal{C}$ can be given a commutative unitary ring structure.*

For $u, v \in R$ take the homomorphism $\varphi : \mathbb{Z} \to R$ defined by $\varphi(1) = v$ and define $u \cdot v = \varphi_r(u)$.

The binary operation $\cdot$ on R has $r_{\mathbb{Z}}(1)$ for its unit since $(r_{\mathbb{Z}})_r$ is the identity on R (thus $u \cdot r_{\mathbb{Z}}(1) = u$ – the other equality $r_{\mathbb{Z}}(1) \cdot u = u$ follows directly from the definition of $\cdot$).

The binary operation $\cdot$ on R is distributive: It suffices to realize that $(f + g)_r = f_r + g_r$ in our notation since then, if $\varphi, \psi : \mathbb{Z} \to R, \varphi(1) = v_1, \psi(1) = v_2$, we have $u \cdot (v_1 + v_2) = (\varphi + \psi)_r(u) = \varphi_r(u) + \psi_r(u) =$

$u \cdot v_1 + u \cdot v_2$. The other equality $(u_1 + u_2) \cdot v$ follows trivially from the fact that φ_r are homomorphisms.

Because of distributivity, the maps $g, h : R \to R$ assigning to u respectively the point $u \cdot v$ or $v \cdot u$, for a fixed $v \in R$, are homomorphisms with $g \circ r_{\mathbb{Z}} = h \circ r_{\mathbb{Z}}$ (indeed, for $n \in \mathbb{Z}$ one has clearly $r_{\mathbb{Z}}(n) \cdot v = nv = v \cdot r_{\mathbb{Z}}(n)$), so that they must coincide on R. Consequently, the binary operation $\cdot$ on R is commutative.

It remains to show that the operation $\cdot$ is associative. Suppose we have $t, u, v \in R$ and $\varphi, \tau : \mathbb{Z} \to R$ with $\varphi(1) = v, \tau(1) = u$. Then $(t \cdot u) \cdot v = \varphi_r(t \cdot u) = \varphi_r(\tau_r(t)) = (\varphi_r \circ \tau)_r(t) = t \cdot (\varphi_r \circ \tau)(1) = t \cdot \varphi_r(u) = t \cdot (u \cdot v)$.

Therefore, $(R, +, \cdot)$ is a commutative unitary ring. If we need to distinguish between the group and the ring structure on R, we shall denote by R the group and by $F(R)$ the ring.

Claim 2. *There is a functor $F : \mathcal{C} \to \mathtt{Mod}_R$ such that $U \circ F$ is the inclusion of $\mathcal{C}$ into* $\mathtt{Ab}$.

For $G \in \mathcal{C}$ we define an external multiplication $R \times G \to G$ as follows: for $g \in G$ take the homomorphism $\varphi : \mathbb{Z} \to G$ with $\varphi(1) = g$ and for $u \in R$ define $u \cdot g = \varphi_r(u)$. The analogous procedure as above shows that every $G \in \mathcal{C}$ with the external multiplication forms an R-module (denoted by $F(G)$). For every homomorphism $f : G \to H$ between members of $\mathcal{C}$ we have $f(u \cdot g) = f(\varphi_r(u)) = (f \circ \varphi)_r(u) = u \cdot f(\varphi(1)) = u \cdot f(g)$ for every $u \in R, g \in G$ — consequently, every group homomorphism $f \in \mathcal{C}(G, H)$ is an R-module homomorphism $F(f) \in \mathtt{Mod}_R(F(G), F(H))$. It is easy to see that $F : \mathcal{C} \to \mathtt{Mod}_R$ is a functor having the required property.

Claim 3. *$F \circ U$ is the identity on* $\mathtt{Mod}_R$.

At first we show that $U(M) \in \mathcal{C}$ for $M \in \mathtt{Mod}_R$. Define a map $f : U(M) \to c(U(M))$ by $f(x) = \widetilde{\widetilde{\varphi_x}}(r_{\mathbb{Z}}(1))$, where $\varphi_x \in \mathtt{Ab}(\mathbb{Z}, U(M))$ with $\varphi_x(1) = x$, $\widetilde{\varphi_x} : F(R) \to M$ is the unique morphism in $\mathtt{Mod}_R$ with $U(\widetilde{\varphi_x}) \circ r_{\mathbb{Z}} = \varphi_x$ (realize that U is adjoint), and $\widetilde{\widetilde{\varphi_x}} : R \to c(U(M))$ is the unique morphism in $\mathtt{Ab}$ with $c_{U(M)} \circ \widetilde{\widetilde{\varphi_x}} = U(\widetilde{\varphi_x})$. The following diagram shows the situation:

$$\begin{tikzcd}
\mathbb{Z} \arrow[r, "r_{\mathbb{Z}}"] \arrow[drr, "\varphi_x"'] & R = U(F(R)) \arrow[dl, "\widetilde{\widetilde{\varphi_x}}"] \arrow[d, "U(\widetilde{\varphi_x})"] & \\
c(U(M)) \arrow[r, "c_{U(M)}"'] & U(M) &
\end{tikzcd}
\qquad
\begin{tikzcd}
F(R) \arrow[d, "\widetilde{\varphi_x}"] \\
M
\end{tikzcd}$$

Since $\widetilde{\varphi_{x+y}} = \widetilde{\varphi_x} + \widetilde{\varphi_y}$ and $c_{U(M)} \circ (\widetilde{\widetilde{\varphi_x}} + \widetilde{\widetilde{\varphi_y}}) = U(\widetilde{\varphi_x} + \widetilde{\varphi_y}) = U(\widetilde{\varphi_{x+y}})$, we have $\widetilde{\widetilde{\varphi_x}} + \widetilde{\widetilde{\varphi_y}} = \widetilde{\widetilde{\varphi_{x+y}}}$ and, thus, f is an $\mathtt{Ab}$-homomorphism.

For each $x \in M$ we have $(c_{U(M)} \circ f)(x) = c_{U(M)}(\widetilde{\widetilde{\varphi_x}}(r_{\mathbb{Z}}(1))) = U(\widetilde{\varphi_x})(1) = \varphi_x(1) = x$ and, therefore $c_{U(M)} \circ f = 1_{U(M)}$. Consequently, $U(M)$ is a retract of $c(U(M))$ in $\mathtt{Ab}$ and must belong to $\mathcal{C}$.

The last step is to show that $U : \mathtt{Mod}_R \to \mathtt{Ab}$ is a full embedding. It suffices to show that U is 1-1 on objects, i.e., that every $G \in \mathcal{C}$ has only one R-module structure. Take $M \in \mathtt{Mod}_R$ with an external multiplication $u{\times}x$; we want to show that it coincides with the multiplication $u \cdot x$ on $F(U(M))$. By definition, $u \cdot x = \varphi_r(u)$, where $\varphi : \mathbb{Z} \to U(M), \varphi(1) = x$. Since $R = U(F(R))$, there is exactly one $h \in \mathtt{Mod}_R(F(R), M)$ with $U(h) \circ r_{\mathbb{Z}} = \varphi$. But there is only one $\mathtt{Ab}$-homomorphism $\varphi_r : R \to U(M)$ having the property $\varphi_r \circ r_{\mathbb{Z}} = \varphi$ and, therefore, $U(h) = \varphi_r$. Consequently, $u{\cdot}x = \varphi_r(u) = U(h)(u) = h(u) = h(u{\cdot}1) = u{\times}h(r_{\mathbb{Z}}(1)) = u{\times}\varphi(1) = u{\times}x$. ■

Remarks:

(1) The first part of the previous proof gives an abstract proof of the known assertion that the forgetful functor $U : \mathtt{Mod}_R \to \mathtt{Ab}$ is both adjoint and coadjoint. Its right adjoint is the functor $\mathtt{Ab} \to \mathtt{Mod}_R$ assigning $R \otimes X$ to X and the left adjoint of U assigns Hom (R, X) to X (see e.g., [9, pp.85, 87]). — the module multiplication $R{\times}$Hom $(R, X) \to$ Hom (R, X) is defined by $(u, f)(v) = f(u \cdot v)$.

(2) The second part of the previous proof (Claims 1 and 2) shows that if $\mathcal{C}$ is reflective in $\mathtt{Ab}$, then the corresponding reflection R of $\mathbb{Z}$ possesses a commutative unitary ring structure such that $\mathcal{C}$ can be regarded as a full subcategory of $\mathtt{Mod}_R$ (it coincides with $\mathtt{Mod}_R$ iff C is coreflective in $\mathtt{Ab}$).

3 Characterization of the Ring R

3.1 Background

Let $\mathcal{A}$ be a subcategory of the category $\mathtt{Ab}$ of Abelian groups. Because of the well-known nice features of $\mathtt{Ab}$ (e.g., completeness, cocompleteness, wellpoweredness, co-wellpoweredness, algebraicness, Abelianness) the following facts are immediate consequences of known general results or easily established:

1. $\mathtt{Ab}$ *has no proper bireflective or bicoreflective subcategory.*

2. $\mathtt{Ab}$ *has no proper reflective subcategory that contains a coseparator (e.g., $\mathbb{Q}/\mathbb{Z}$) of* $\mathtt{Ab}$.

And dually: $\mathtt{Ab}$ *has no proper coreflective subcategory that contains a separator (e.g., $\mathbb{Z}$) of* $\mathtt{Ab}$.

(Immediate from the previous item since such subcategories are bireflective resp. bicoreflective.)

3. *$\mathcal{A}$ is epireflective in* $\mathtt{Ab}$ *iff $\mathcal{A}$ is closed under the formation of products and of subgroups.*

And dually: *$\mathcal{A}$ is monocoreflective in* $\mathtt{Ab}$ *iff $\mathcal{A}$ is closed under the formation of direct sums (= coproducts) and homomorphic images (= quotients).*

4. *If $\mathcal{A}$ is reflective and closed under homomorphic images in* $\mathtt{Ab}$ *, then*

$\mathcal{A}$ is epireflective.
And dually: *If $\mathcal{A}$ is coreflective and closed under subgroups in* $\mathtt{Ab}$ *, then $\mathcal{A}$ is monocoreflective.*
In fact, these results hold for Abelian categories.

3.2 Epireflective and Monocoreflective

Since simultaneously epireflective and monocoreflective subcategories of $\mathtt{Ab}$ coincide with classes of Abelian groups that are closed under the formation of products, direct sums, subgroups, and homomorphic images, they coincide with varieties in $\mathtt{Ab}$, so that their characterization is known (see e.g., [2, p.93]). Nevertheless, we shall show how to obtain the characterization by categorical methods and get some other relations, too.

We shall use the usual notation for $n \in \mathbb{N} \cup \{0\}$: $\mathbb{Z}(\mathbf{n}) = \mathbb{Z}/\mathbf{n}\mathbb{Z}$. By $\mathtt{Tor}(n)$, the subcategory of $\mathtt{Ab}$ consisting of all *n-torsion groups*, i.e., all groups A with $n \cdot A = 0$ will be denoted (called n-bounded in [2]). Realize that except the usual groups $\mathbb{Z}(n)$ for $n \geq 2$ we have two extreme cases, namely $n = 0$ and $n = 1$:

$$\mathbb{Z}(0) = \mathbb{Z}, \mathtt{Tor}(0) = \mathtt{Ab} \qquad \mathbb{Z}(1) = 0, \mathtt{Tor}(1) = \{0\}.$$

Since $\mathtt{Tor}(n)$ is closed under the formation of products, direct sums, subgroups, and homomorphic images, we get from the previous fact 3:

Proposition 3.1 *Each* $\mathtt{Tor}(n)$ *is simultaneously epireflective and monocoreflective in* $\mathtt{Ab}$.

Proposition 3.2
1. *For each n, $\mathbb{Z}(n)$ is a projective separator for* $\mathtt{Tor}(n)$.
2. *For each $n \neq 0$, $\mathbb{Z}(n)$ is an injective coseparator for* $\mathtt{Tor}(n)$.

Proof.
1. Straightforward and known.
2. This is trivially true for $n = 1$. For $n \geq 2$ our result follows easily from the known representation of n-torsion groups as direct sums of $\mathbb{Z}(p^i)$ for divisors p^i of n. ■

Proposition 3.3
1. *For each n,* $\mathtt{Tor}(n)$ *has no proper coreflective subcategory that contains $\mathbb{Z}(n)$.*
2. *For $n \neq 0$,* $\mathtt{Tor}(n)$ *has no proper reflective subcategory that contains $\mathbb{Z}(n)$.*

Proof. The mentioned subcategories are bireflective or bicoreflective in $\mathtt{Tor}(n)$. ■

Proposition 3.4
1. *For each n,* $\mathtt{Tor}(n)$ *is simultaneously:*
   (a) *the coreflective hull of* $\mathbb{Z}(n)$ *in* $\mathtt{Ab}$,
   (b) *the monocoreflective hull of* $\mathbb{Z}(n)$ *in* $\mathtt{Ab}$.
2. *For each $n \neq 0$,* $\mathtt{Tor}(n)$ *is simultaneously:*
   (a) *the reflective hull of* $\mathbb{Z}(n)$ *in* $\mathtt{Ab}$,
   (b) *the epireflective hull of* $\mathbb{Z}(n)$ *in* $\mathtt{Ab}$.

Proof. Immediate from Propositions 3.1 and 3.3. ∎

Proposition 3.5 *For each epireflective subcategory $\mathcal{A}$ of* $\mathtt{Ab}$ *there exists n with $\mathbb{Z}(n) \in \mathcal{A} \subset \mathtt{Tor}(n)$.*

Proof. Let $r\colon \mathbb{Z} \longrightarrow A_{\mathbb{Z}}$ be an $\mathcal{A}$-reflection arrow for $\mathbb{Z}$. If r is injective, then $\mathbb{Z}$ belongs to $\mathcal{A}$, and thus $\mathbb{Z}(0) = \mathbb{Z} \in \mathcal{A} \subset \mathtt{Ab} = \mathtt{Tor}(0)$.

If r is not injective let n be the smallest positive integer with $r(n) = 0$. Then there exists a factorization

$$\mathbb{Z} \xrightarrow{r} A_{\mathbb{Z}} = \mathbb{Z} \xrightarrow{e} \mathbb{Z}(n) \xrightarrow{m} A_{\mathbb{Z}}$$

with e epi and m mono. Consequently, $\mathbb{Z}(n)$ belongs to $\mathcal{A}$ since m is also epi, and $\mathcal{A} \subset \mathtt{Tor}(n)$, since each map $f\colon \mathbb{Z} \longrightarrow A$ from $\mathbb{Z}$ into an $\mathcal{A}$-object factors through r and thus through e.

Thus $\mathbb{Z}(n) \in \mathcal{A} \subset \mathtt{Tor}(n)$. ∎

Theorem 3.6 *Equivalent are:*
1. *$\mathcal{A}$ is epireflective and monocoreflective in* $\mathtt{Ab}$,
2. *$\mathcal{A}$ is reflective and monocoreflective in* $\mathtt{Ab}$,
3. *$\mathcal{A}$ is epireflective and coreflective in* $\mathtt{Ab}$,
4. *$\mathcal{A}$ is a subvariety of* $\mathtt{Ab}$,
5. *there exist (a unique) $n \in \mathbb{N}$ with $\mathcal{A} = \mathtt{Tor}(n)$.*

Proof. (1) $\Leftrightarrow$ (4) Immediate in view of the characterization of epireflectivity and monocoreflectivity in $\mathtt{Ab}$ and the fact that in $\mathtt{Ab}$ direct sums are subgroups of products.
(1) $\Rightarrow$ (2) Obvious.
(2) $\Rightarrow$ (3) Immediate via the fact 4 from the preceding section.
(3) $\Rightarrow$ (5) By Proposition 3.5 there exists $n \in \mathbb{N}$ with $\mathbb{Z}(n) \in \mathcal{A} \subset \mathtt{Tor}(n)$. Thus $\mathcal{A} = \mathtt{Tor}(n)$ by Proposition 3.4 (1a). Uniqueness of n follows immediately from the fact that $n < m$ implies $\mathbb{Z}_m \notin \mathtt{Tor}(n)$.
(5) $\Rightarrow$ (1) Immediate by Proposition 3.1. ∎

3.3 Reflective and Coreflective

We know from Theorem 2.1 that every full subcategory of $\mathtt{Ab}$ that is simultaneously reflective and coreflective coincides with a category $\mathtt{Mod}_R$ that is canonically fully embedded into $\mathtt{Ab}$. The previous subsection described the cases when R is finite (or equivalently, torsion); the resulting categories of n-torsion groups were both epireflective and monocoreflective. If there is any other both reflective and coreflective subcategory of $\mathtt{Ab}$, it can be neither epireflective nor monocoreflective.

We shall now investigate those infinite (or nontorsion) commutative unitary rings R such that the category of the corresponding R-modules is a full subcategory of $\mathtt{Ab}$, i.e., every group homomorphism between two R-modules is also an R-module homomorphism. So in this subsection, we shall suppose that $\mathcal{C}$ is *a reflective and coreflective full subcategory of* $\mathtt{Ab}$ *containing a nontorsion group.*

A significant role in characterizing $\mathcal{C}$ will be played by the groups $\mathbb{Z}(p^k)$, where p is a prime and k is a natural number or 0 or ∞. We shall denote by $\mathcal{P}$ the set of all primes and by $\mathbb{N}^*$ the set of natural numbers extended by 0 and ∞.

For $p \in \mathcal{P}$ let

$$n_p = \sup\{k \in \mathbb{N}^* : \mathbb{Z}(p^k) \in \mathcal{C}\},$$

thus $n_p \in \mathbb{N}^*$. We shall see later that the sequence $\{n_p\}$ will characterize the subcategory $\mathcal{C}$.

We shall often use the following partition $\{S_0, S_{\mathbb{N}}, S_\infty\}$ of $\mathcal{P}$:

$$S_0 = \{p \in \mathcal{P} : n_p = 0\}, \quad S_{\mathbb{N}} = \{p \in \mathcal{P} : n_p \in \mathbb{N}\} \quad S_\infty = \{p \in \mathcal{P} : n_p = \infty\}.$$

By $r_{\mathbb{Z}} : \mathbb{Z} \to R$ we denote a reflection of $\mathbb{Z}$ in $\mathcal{C}$. Realize that our assumption on $\mathcal{C}$ ($\mathcal{C}$ contains a nontorsion group) implies that $r_{\mathbb{Z}}$ is a monomorphism. Nevertheless, most of the next results are valid even when $r_{\mathbb{Z}}$ is not a monomorphism (i.e., when $\mathcal{C} = \mathtt{Tor}(n)$).

Observations:

1. *Let* $G \in \mathcal{C}$ *and* G_0 *be a subgroup of* G*. Then* $G_0 \in \mathcal{C}$ *iff* $G/G_0 \in \mathcal{C}$*.* Indeed, we have the following commutative diagram

$$\begin{tikzcd}
G_0 \arrow[rr, "i"] \arrow[dr] & & G \arrow[rr, "j"] \arrow[dr, "j_c"'] & & G/G_0 \\
& r(G_0) \arrow[ur, "i_r"'] & & c(G/G_0) \arrow[ur, "c"'] &
\end{tikzcd}$$

Suppose first that $G_0 \in \mathcal{C}$. Since $j \circ i = 0$ we have $j_c \circ i = 0$ and, thus, there is a homomorphism $\varphi : G/G_0 \to c(G/G_0)$ with $\varphi \circ c = 1$, which means that the epimorphism c is a monomorphism, thus an isomorphism. If $G/G_0 \in \mathcal{C}$ then $j \circ i_r = 0$ and, hence, $i_r(r(G_0)) \subset G_0$. Thus $r_{G_0} \circ i_r = 1$ and the monomorphism r_{G_0} is an epimorphism, thus an isomorphism. (It is easy to prove our observation also by means of elementary facts from module theory.)

2. $\mathbb{Z}(p^k) \in \mathcal{C}$ *for every* $0 \le k \le n_p$. Indeed, for some $0 < k \le n_p$ take the monomorphism $f : \mathbb{Z}(p^k) \to \mathbb{Z}(p^{n_p})$ generated by $f(1) = p^{n_p - k}$ if $n_p < \infty$

and by $f(1) = p^{-k}$ otherwise, and take its extension $f_r : r(\mathbb{Z}(p^k)) \to \mathbb{Z}(p^{n_p})$. Since $p^k \cdot f = 0$, we have $p^k \cdot f_r = 0$ because of uniqueness of extensions. Consequently, the image of f_r is contained in the image of f and, thus, $\mathbb{Z}(p^k)$ is a retract of $r(\mathbb{Z}(p^k))$. Therefore, it belongs to $\mathcal{C}$.

3. *$\mathbb{Z}(p) \in \mathcal{C}$ iff $p \cdot R \neq R$ (i.e., R is not divisible by p).* Indeed, if $\mathbb{Z}(p) \in \mathcal{C}$ then the canonical map $\mathbb{Z} \to \mathbb{Z}(p)$ transfers via $r_{\mathbb{Z}}$ and, hence, $r_{\mathbb{Z}}(1)$ is not divisible by p (otherwise its image 1 in $\mathbb{Z}(p)$ is divisible by p). Conversely, if $p \cdot R \neq R$ then $R/(p \cdot R)$ is a nontrivial direct sum of $\mathbb{Z}(p)$'s and, by 1, it suffices to realize that $p \cdot R \in \mathcal{C}$. (This item 3 is not valid if we change p to p^k for $k > 1$.)

4. *For a prime $p, p \in S_0$ iff R is divisible by p (i.e., iff $r_{\mathbb{Z}}(1)$ is divisible by p).* The first equivalence follows from the preceding items 2 and 3. If $r_{\mathbb{Z}}(1)$ is divisible by p (say, $py = r_{\mathbb{Z}}(1)$) and $x \in R$, take the homomorphism $f : \mathbb{Z} \to R$ defined by $f(1) = x$; then $pf_r(y) = x$.

5. *For a prime p, if $p \in S_{\mathbb{N}}$ then $pu = 0$ for some nonzero $u \in R$* (the converse implication will follow from Claim 1 of Proposition 3.8 and, thus, the division by $p \in S_0$ from the item 4 is unique.) The quotient $R/(p^k R)$ is isomorphic to a direct sum of some $\mathbb{Z}(p^i)$'s. Thus, if $\mathbb{Z}(p^k) \notin \mathcal{C}$, all i's in the sum must be smaller than k and, hence, there is some $y \in R$ and $i < k$ with $p^i r_{\mathbb{Z}}(1) = p^k y$. Since $p \notin S_0$, we have $r_{\mathbb{Z}}(1) - p^{k-i} y \neq 0$ and there exists a smallest natural number j with $p^j(r_{\mathbb{Z}}(1) - p^{k-i}y) = 0$ and then $pu = 0$ for $0 \neq u = p^{j-1}(r_{\mathbb{Z}}(1) - p^{k-i}y)$.

6. *R/tR is divisible by all $p \in S_0 \cup S_{\mathbb{N}}$.* Because of the item 4, it remains to prove that $p(R/tR) = R/tR$ for $p \in S_{\mathbb{N}}$. If the equality does not hold, we have the diagram

$$\begin{tikzcd}
R/tR \arrow[r, "f'"] & (R/tR)/(p \cdot R/tR) \arrow[dr, "\varphi"] & \\
R \arrow[u, "f"] \arrow[r, "g"] & R/(p \cdot R) \arrow[u] \arrow[r, "\psi"] & \mathbb{Z}(p) \\
\mathbb{Z} \arrow[u, "r_{\mathbb{Z}}"] \arrow[r, "h"] & \mathbb{Z}/(p \cdot \mathbb{Z}) \arrow[u] \arrow[ur, "\tau"] &
\end{tikzcd}$$

where f, f', g, h are the factor maps and τ, φ, ψ are epimorphisms of corresponding G/pG onto $\mathbb{Z}(p)$. Since $g(r_{\mathbb{Z}}(1)) \neq 0$ there exists a ψ with $\psi(g(r_{\mathbb{Z}}(1))) = 1$. Similarly we can find φ with $\varphi(f'(f(r_{\mathbb{Z}}(1)))) = 1$ and τ with $\tau(h(1)) = 1$. It is now sufficient to realize that there exists $x \in tR$ not divisible by p (indeed, by the item 5, there is some $y \neq 0$ with $py = 0$, and such y cannot be divisible by all p^k since otherwise we could embed $\mathbb{Z}(p^\infty)$ into R and p would belong to S_∞). Then $f(x) = 0$ but one can find ψ having the above property and, moreover, $\psi(g(x)) \neq 0$. We have obtained two different homomorphisms $R \to \mathbb{Z}(p)$ extending $\tau \circ h$ – a contradiction.

Proposition 3.7 *Under the condition that $r_{\mathbb{Z}}$ is a monomorphism, the group of rationals $\mathbb{Q}$ belongs to $\mathcal{C}$.*

Proof. If $S_\infty = \emptyset$, then the nonzero group R/tR is divisible, as follows from our Observation 6, and belongs to $\mathcal{C}$. Since it is torsion-free, it is isomorphic to a direct sum of $\mathbb{Q}$'s. Consequently, $\mathbb{Q} \in \mathcal{C}$ as a retract of the sum.

If $p \in S_\infty$, then $mx \in r_{\mathbb{Z}}(\mathbb{Z})$ for every $x \in R$ and some $m \in \mathbb{N}$ (indeed, otherwise we can construct a nonzero $f : R \to \mathbb{Z}(p^\infty)$ with $f \circ r_{\mathbb{Z}} = 0$, namely an extension with $f(r_{\mathbb{Z}}(n) + mx) = m \cdot (1/p)$). Take now a coreflection $c_{\mathbb{Q}} : c\mathbb{Q} \to \mathbb{Q}$ and suppose that $c_{\mathbb{Q}}(y) = 0$. For the homomorphism $\varphi : \mathbb{Z} \to c_{\mathbb{Q}}\mathbb{Q}$ with $\varphi(1) = y$ we have $c_{\mathbb{Q}} \circ \varphi_r = 0$ (indeed, for $x \in R$ we have $mx = r_{\mathbb{Z}}(n)$ for some $m \in \mathbb{N}$ so that $m\, c_{\mathbb{Q}}(\varphi_r(x)) = c_{\mathbb{Q}}(\varphi_r(mx)) = c_{\mathbb{Q}}(\varphi_r(r_{\mathbb{Z}}(n))) = c_{\mathbb{Q}}(\varphi((n \cdot 1)) = n\, c_{\mathbb{Q}}(y) = 0)$. Thus $\varphi_r = 0$ and $\varphi = 0$ and $y = 0$, which means that $c_{\mathbb{Q}}$ is a monomorphism. Because it is an epimorphism ($\mathbb{Q}$ is injective), it is an isomorphism, and $\mathbb{Q} \in \mathcal{C}$. ■

We shall now describe the torsion part of R.

Proposition 3.8 *The torsion subgroup tR of R is isomorphic to the direct sum $\sum_{p \in S_{\mathbb{N}}} \mathbb{Z}(p^{n_p})$.*

Proof. By [2, Th.8.4], tR is the direct sum of p-groups A_p belonging to different primes p. Since $tR \in \mathcal{C}$, every A_p belongs to $\mathcal{C}$. The next claim and the fact that $A_p/p^k A_p$ belongs to $\mathcal{C}$ imply that every A_p is n_p-torsion.

Claim 1. *If $x \neq 0$ is a torsion element of R and a prime p divides $ord(x)$, then $p \in S_{\mathbb{N}}$.*

Let $py = 0$ in R and $y \neq 0$. If $p \in S_0$ then, by Observation 3, we may divide y by any p^k, which implies that $\mathbb{Z}(p^\infty)$ embeds into R and thus, it is a retract of R (as an injective group) and, therefore, $p \in S_\infty$ — a contradiction. If $p \in S_\infty$, then the zero homomorphism $\mathbb{Z} \to \mathbb{Z}(p^\infty)$ extends to a nonzero homomorphism on R (namely, we may first extend the zero map to the smallest subgroup of R containing $r_{\mathbb{Z}}(\mathbb{Z}) \cup \{y\}$ by assigning $1/p$ to y, and then to whole R). That contradicts the reflectivity property of R. Consequently, $p \in S_{\mathbb{N}}$.

Now we know that A_p is nonzero for $p \in S_{\mathbb{N}}$ only, and for those p it is a direct sum of groups $\mathbb{Z}(p^i)$ for some numbers (they may repeat) $0 < i \leq n_p$. Our proof will be complete after proving the following Claim.

Claim 2. *If $p \in S_{\mathbb{N}}$ then A_p is isomorphic to $\mathbb{Z}(p^{n_p})$.*

If there are two copies of some $\mathbb{Z}(p^i)$ and $\mathbb{Z}(p^j)$ for $0 < i \leq j$ in the direct sum representation of A_p, it is not difficult to construct a nonzero homomorphism on R into $\mathbb{Z}(p^j)$ that is zero on $r_{\mathbb{Z}}(\mathbb{Z})$, which contradicts the reflectivity of R. So, there is exactly one nonzero i in the above direct sum. It remains to show that $i = n_p$. In case $i < n_p$, the epimorphism

$f : \mathbb{Z} \to \mathbb{Z}(p^{n_p})$ generated by $f(1) = 1$ extends to R and f_r factorizes via $R/(p^{n_p} \cdot R)$. Since R/tR is divisible by p, every $x \in R$ can be expressed as $p^{n_p}y + z$, where $z \in tR$, and hence, $f_r(x) = f_r(z)$. But f_r must be 0 on every $A_q, q \neq p$ and cannot be an epimorphism on A_p if $i < n_p$. Thus $i = n_p$. Claim 2 is proved. ■

Corollary 3.9 *If $S_{\mathbb{N}}$ is finite and $n = \Pi_{S_{\mathbb{N}}} p^{n_p}$, then the torsion part tR of R is a retract of R and, thus, R is isomorphic to $(n \cdot R) \times tR$.*

Proof. Since R/tR is divisible by n, every $x \in R$ can be written uniquely (realize that $(n \cdot R) \cap tR = 0$) as $ny + z$ for some $y \in R$ and $z \in tR$. ■

Corollary 3.10 *For every $x \in R$ there exists some $m \in \mathbb{N}$ with $mx \in r_{\mathbb{Z}}(\mathbb{Z})$. If m is such a smallest number, then no $p \in S_\infty$ divides m.*

Proof. Our first assertion is equivalent to the uniqueness of extensions of homomorphisms $\mathbb{Z} \to \mathbb{Q}$ to R.

Let m be the smallest natural number with $mx \in r_{\mathbb{Z}}(\mathbb{Z})$. If $mx = 0$ then $x \in tR$ and we know from Proposition 3.8 that m is divided only by $p \in S_{\mathbb{N}}$. Assume that $x \notin tR, mx = r_{\mathbb{Z}}(n)$ and $p \mid m$, say $m = kp$. If $n = pl$ then $p(kx - r_{\mathbb{Z}}(l)) = 0$, which means $p \in S_{\mathbb{N}}$. If $p \nmid n$ and $ap + bn = 1$ for some $a, b \in \mathbb{Z}$, then $p(r_{\mathbb{Z}}(a) + bk\,x) = r_{\mathbb{Z}}(1)$ and $p \in S_0$. ■

We have a nice representation of the ring R in case R is torsion (see the preceding subsection) and in case R has a finite torsion part (see Corollary 3.9). It remains to find a representation in case R has an infinite torsion part, i.e., in case the set $S_{\mathbb{N}}$ is infinite. Take the ring product

$$G_C = \Pi_{p \in S_{\mathbb{N}}} \mathbb{Z}(p^{n_p}),$$

and denote by H_C its subgroup

$$H_C = \{x \in G_C : m \cdot x = n \cdot 1_{G_C} \text{ for some } m \in \mathbb{N}, n \in \mathbb{Z} \text{ with } p \nmid m \text{ for } p \in S_\infty\}$$

(where 1_{G_C} is the unit of the ring G_C). Realize that H_C is a subring of G_C and the torsion part tH_C coincides with $\sum_{S_{\mathbb{N}}} \mathbb{Z}(p^{n_p})$.

There is a homomorphism $i : R \to G_C$ extending the canonical homomorphism $\mathbb{Z} \to G_C$ (where 1 is sent to 1 in every $\mathbb{Z}(p^{n_p})$). We shall now characterize the image of i.

Proposition 3.11 *If $S_{\mathbb{N}}$ is infinite, then the homomorphism i is an isomorphism between R and H_C.*

Proof. Because of Corollary 3.10 we have $i(R) \subset H_C$.

The map i is a monomorphism: Proposition 3.8 asserts that i is a monomorphism on the torsion part tR of R. Take $x \in R$ with infinite order. Then there are some $n, m \in \mathbb{Z}$ such that $m \cdot x = r_{\mathbb{Z}}(n)$. There must be $p \in S_{\mathbb{N}}$ such that $p \nmid n$ and, thus, $i(x)$ has the p-th coordinate nonzero.

The map i is an epimorphism: Clearly, i maps the torsion part of R onto that of $G_{\mathcal{C}}$. So, take some $x \in H_{\mathcal{C}}$ with infinite order. We may assume that $mx = \{n\}$ for some $m, n \in \mathbb{N}$ such that no $p \in S_\infty$ divides m. Consequently, R/tR is divisible by m and $r_{\mathbb{Z}}(n) = mu + z$ for some $u \in R, z \in tR$, thus $\{n\} = m\,i(u) + i(z)$ and $m(x - i(u)) = i(z)$. The last equality means that $x - i(u) \in t(H_{\mathcal{C}})$ so that $x - i(u) = i(v)$ for some $v \in tR$ and, hence, $x = i(u + v)$. The proof is complete. ∎

We shall now summarize the preceding representation results. By $\mathbb{Q}_A$, for $A \subset \mathcal{P}$, we denote the subring of $\mathbb{Q}$ composed of all quotients with denominators not divisible by $p \in A$. If H is a subgroup of a group G and $A \subset \mathcal{P}$, then the group $\{x \in G$: there is some $m \in \mathbb{N}$ coprime to every $p \in A$ such that $mx \in H\}$ is called the *weak* $(\mathcal{P} \setminus A)$*-essential hull* of H in G. By Z we shall denote the i-image of $\mathbb{Z}$ in $G_{\mathcal{C}}$.

Theorem 3.12 *The commutative unitary rings R such that the forgetful functor $U : \mathtt{Mod}_R \to \mathtt{Ab}$ is a full embedding are exactly those isomorphic to one of the following rings:*

1. If R is a torsion group, then $R \cong \mathbb{Z}_n$ for some $n \in \mathbb{N}$. In this case $S_\infty = \emptyset, |S_{\mathbb{N}}| < \omega$ ($S_{\mathbb{N}} = \{p : p \mid n\}, n_p = \max\{k : p^k \mid n\}$), $U(\mathtt{Mod}_R) = \mathtt{Tor}(n)$ is the class of n-torsion groups.

2. If R is not torsion and its torsion part is finite, then $R \cong \mathbb{Q}_A \times \mathbb{Z}_n$, where no $p \in A$ divides n. In this case $S_\infty = A, S_{\mathbb{N}}$ is finite and described as in the case 1. Every object of $U(\mathtt{Mod}_R)$ is isomorphic to $G \times H$, where G is a group uniquely divisible by every $p \in \mathcal{P} \setminus A$ and $H \in \mathcal{Z}_n$.

3. If R is not torsion and its torsion part is infinite, then there are disjoint subsets A, B of $\mathcal{P}$, B infinite, and numbers $k_p \in \mathbb{N}$ for $p \in B$ such that R is isomorphic to the weak $(\mathcal{P} \setminus A)$-essential hull of Z in $\Pi_{p \in B}\mathbb{Z}(p^{n_p})$. In this case $S_\infty = A, S_{\mathbb{N}} = B$ and $n_p = k_p$ for $p \in B$.

Some interesting special cases of Theorem 4:
$\mathcal{C} = \mathtt{Ab}$ iff $S_0 = S_{\mathbb{N}} = \emptyset$.
$\mathcal{C} = \{0\}$ iff $S_\infty = S_{\mathbb{N}} = \emptyset$ and $\mathcal{C} \subset \mathtt{Tor}$.
$\mathcal{C} = \mathcal{Q}$ iff $S_\infty = S_{\mathbb{N}} = \emptyset$ and $\mathcal{C} \not\subset \mathtt{Tor}$.
$\mathcal{C} = \mathcal{Z}_2$ iff $S_\infty = \emptyset, S_{\mathbb{N}} = \{2\}, n_2 = 1$ and $\mathcal{C} \subset \mathtt{Tor}$.
$\mathcal{C}$ is generated by $\mathbb{Q}$ and $\mathcal{Z}_2$ iff $S_\infty = \emptyset, S_{\mathbb{N}} = \{2\}, n_2 = 1$ and $\mathcal{C} \not\subset \mathtt{Tor}$.
$\mathcal{C}$ is generated by $\mathbb{Q}_{\{2\}}$ iff $S_\infty = \{2\}, S_{\mathbb{N}} = \emptyset$.

In case $S_\infty = \emptyset$, the same choices of $S_0, S_{\mathbb{N}}$ and of n_p determine two different rings — one is torsion and the other nontorsion. If $S_\infty \neq \emptyset$ then R is always nontorsion.

Theorem 3.13 *Subcategories of* $\mathtt{Ab}$ *being simultaneously reflective and coreflective in* $\mathtt{Ab}$ *and containing a torsion-free group are exactly the following subcategories $\mathcal{C}$ characterized by a partition $\{S_0, S_{\mathbb{N}}, S_\infty\}$ of primes and by* *a sequence* $\{p^{n_p} : p \in S_{\mathbb{N}}\}$ *with* $n_p \in \mathbb{N}$*: an Abelian group* X *belongs to* $\mathcal{C}$ *if and only if*

1. *if* $x \in X$ *has a finite order* m*, then the prime decomposition of* m *contains* p^k *only, with* $p \in S_{\mathbb{N}} \cup S_{\infty}, k \leq n_p$*;*

2. *every* $x \in X$ *is uniquely divisible by* $p \in S_0$*;*

3. *every* $x \in X$ *is divisible modulo the torsion part of* X *by* $p \in S_{\mathbb{N}}$*.*

4 A Special Case of Rings without Torsion Part

4.1 General Cases

The cases from Theorem 3.12 when $S_{\mathbb{N}} = \emptyset$ (i.e., R is torsion-free) can serve as interesting simple examples of various (weakly) reflective or coreflective subcategories or hulls.

Definition 4.1 Let P be a set of prime numbers.

(1) $\mathbb{Q}(P)$ is the subring of the field $\mathbb{Q}$ generated by the set $\{\frac{1}{p} \mid p \in P\}$. $\mathbb{Q}(P)$ will also be regarded (forgetting multiplication) as an object of $\mathtt{Ab}$.

(2) $\mathcal{Q}(P)$ is the subcategory of $\mathtt{Ab}$, consisting of (the additive parts of) all $\mathbb{Q}(P)$-*modules.*

(3) $\mathtt{Torf}(P)$ is the subcategory of $\mathtt{Ab}$, consisting of all P-*torsion-free groups*, i.e., all groups satisfying the implication $p \cdot a = 0 \Rightarrow a = 0$ for each p in P (equivalently, the map $f_p \colon A \longrightarrow A$, defined by $f_p(a) = p \cdot a$, is injective for each p in P).

(4) $\mathtt{Div}(P)$ is the subcategory of $\mathtt{Ab}$, consisting of all P-*divisible groups*, i.e., all groups G for which each of its elements is divisible by every $p \in P$ (equivalently, the map f_p from (3) is surjective for each p in P).

We know from Theorem 2.1 that the forgetful functor $U : \mathtt{Mod}_{\mathbb{Q}(P)} \to \mathcal{Q}(P)$ is a concrete isomorphism.

Remark: 4.2 Clearly:

(1) $\mathcal{Q}(P) = \mathtt{Torf}(P) \cap \mathtt{Div}(P)$ consists of all groups G for which each of its elements is uniquely divisible by every $p \in P$ (equivalently, the map $f_p : G \to G$ from (3) is an isomorphism for each p in P).

(2) If $P \neq \emptyset$ (equivalently: if $\mathcal{Q}(P) \neq \mathtt{Ab}$) then $\mathcal{Q}(P)$ is closed neither under the formation of subgroups nor under the formation of homomorphic images. Thus $\mathcal{Q}(P)$ is neither epireflective nor monocoreflective in $\mathtt{Ab}$.

Next we investigate the two constituents, namely P-divisible groups and P-torsion-free groups separately. It turns out that their features are kind of dual to each other.

Lemma 4.3 *Let P be a set of prime numbers and let $Mon(P)$ be the submonoid of the multiplicative monoid of natural numbers, generated by P. If A is a subgroup of some group B such that for each $b \in B$ there exists $m \in Mon(P)$ with $(m \cdot b) \in A$, then each homomorphism $A \xrightarrow{f} D$ into a P-divisible group D can be extended to a homomorphism $B \xrightarrow{g} D$.*

Proof. By Zorn's Lemma the set of all groups C with $A \subset C \subset B$ that allow a homomorphic extension $C \longrightarrow D$ of f has a maximal element.

Let C be such a maximal element and let $g\colon C \longrightarrow D$ be a homomorphic extension of f. It suffices to show that $C = B$. Assume, on the contrary, that there is some b in $B \backslash C$, and let E be the subgroup of B, generated by $C \cup \{b\}$. Then $I = \{n \in \mathbb{Z} \mid n \cdot b \in C\}$ is an ideal in $\mathbb{Z}$, hence of the form $I = n_0 \cdot \mathbb{Z}$ for some $n_0 \in \mathbb{N}$. Since there exists some $m \in M(P)$ with $m \in I$, and since $M(P)$ is closed under the formation of divisors, we conclude that $n_0 \in M(P)$. Thus there exists some $d_0 \in D$ with $n_0 \cdot d_0 = g(n_0 \cdot b)$. Define a map $h\colon E \longrightarrow D$ by $h(c+nb) = g(b) + n \cdot d_0$, where $c \in C$ and $n \in \mathbb{Z}$. That h is well defined (and hence a homomorphism) results from the following computation:

If $c + n \cdot b = \bar{c} + \bar{n} \cdot b$, then $(n - \bar{n}) \cdot b = (\bar{c} - c) \in C$. Thus there exists some $k \in \mathbb{Z}$ with $(n - \bar{n}) = k \cdot n_0$. Consequently: $g(\bar{c}) - g(c) = g(\bar{c} - c) = g\big((n - \bar{n}) \cdot b\big) = g(k \cdot n_0 \cdot b) = k \cdot g(n_0 \cdot b) = k \cdot n_0 \cdot d_0 = (n - \bar{n}) \cdot d_0$. Thus: $g(\bar{c}) + \bar{n} \cdot d_0 = g(c) + n \cdot d_0$. Thus $h\colon E \longrightarrow D$ is a homomorphic extension of g and hence of f, contradicting the maximality of C. Hence $C = B$ and $g\colon B \longrightarrow D$ is a homomorphic extension of f. ■

Theorem 4.4 *For each set P of primes the following hold:*

(1) $\mathtt{Div}(P)$ *is monocoreflective and closed under products in* $\mathtt{Ab}$,

(2) $\mathtt{Div}(P)$ *is almost (mono) reflective, but for $P \neq \emptyset$ not reflective in* $\mathtt{Ab}$.

Proof. (1) Immediate, since $\mathtt{Div}(P)$ is closed under the formation of products, direct sums, and homomorphic images.

(2) Let A be a group, and let $A \overset{m_A}{\hookrightarrow} DA$ be an injective (= divisible) hull of A in $\mathtt{Ab}$. Let $M(P)$ be the submonoid of the multiplicative monoid of natural numbers, generated by P. Consider the subgroup $D_P A$ of DA, consisting of all elements $x \in DA$ such that there exists some $m \in M(P)$ with $m \cdot x \in A$. Then $D_P A$ is P-divisible, and thus by Lemma 4.3 a weak $\mathtt{Div}(P)$-reflection. Since $\mathtt{Div}(P)$ is also closed under the formation of retracts, it is almost reflective in $\mathtt{Ab}$.

That $\mathtt{Div}(P)$, for $P \neq \emptyset$, is not reflective in $\mathtt{Ab}$ follows from the fact that $\mathtt{Div}(P)$ is not closed under the formation of equalizers in $\mathtt{Ab}$. Consider, e.g., the natural quotient map $\mathbb{Q} \xrightarrow{f} \mathbb{Q}/\mathbb{Z}$ and the zero map $\mathbb{Q} \xrightarrow{g} \mathbb{Q}/\mathbb{Z}$. Then the inclusion map $\mathbb{Z} \hookrightarrow \mathbb{Q}$ is an equalizer of f and g, but whereas $\mathbb{Q}$ and $\mathbb{Q}/\mathbb{Z}$ belong to $\mathtt{Div}(P)$, the group $\mathbb{Z}$ (for $P \neq \emptyset$) fails to do so. ■

Lemma 4.5 *For each P, the injective hull of a P-torsion-free group is P-torsion-free.*

Proof. Let $A \overset{m_A}{\hookrightarrow} DA$ be an injective hull of a P-torsion-free group A. For $p \in P$ consider the map $f_p\colon DA \longrightarrow DA$, defined by $f_p(x) = p \cdot x$. Then the composition $A \xrightarrow{f_p \circ m_A} DA$ restricts to a map $A \xrightarrow{g} A$, defined by $g(a) = p \cdot a$. Since A is P-torsion-free, g and hence $f_p \circ m_A$ are monomorphisms. Since m_A is – by definition – an essential monomorphism, f_p has to be a monomorphism. Thus DA is P-torsion-free. ■

Corollary 4.6 *For each P, $\mathtt{Div}(P)$ is the monocoreflective hull of $\mathbb{Q}(P)$.*

Proof. By Lemma 4.3 the inclusion map $\mathbb{Z} \hookrightarrow \mathbb{Q}(P)$ is a weak $\mathtt{Div}(P)$-reflection. Since for every group A, the sink $(hom(\mathbb{Z}, A), A)$ is an epi-sink, this implies that for every P-divisible group A, the sink $(hom(\mathbb{Q}(P)), A)$ is an epi-sink. Thus A is a quotient of a coproduct of $\mathbb{Q}(P)$'s. ■

Theorem 4.7 *For each set P of primes the following hold:*

(1) $\mathtt{Torf}(P)$ is epireflective (and thus closed under coproducts) in $\mathtt{Ab}$,

(2) $\mathtt{Torf}(P)$ is almost coreflective, but for $P \neq \emptyset$ not coreflective in $\mathtt{Ab}$.

Proof. (1) Immediate, since $\mathtt{Torf}(P)$ is closed under the formation of products and subgroups, hence also of direct sums.

(2) Denote by $X \xrightarrow{m_X} DX$ an injective hull of a group X. Now start with a group A, and let $CA \xrightarrow{c_A} DA$ be a $\mathcal{Q}(P)$-coreflection arrow for DA. Form a pullback diagram

$$\begin{tikzcd} \bar{A} \arrow[r, hook, "\bar{m}_A"] \arrow[d, "\bar{c}_A"'] & CA \arrow[d, "c_A"] \\ A \arrow[r, hook, "m_A"'] & DA \end{tikzcd}$$

Then $\bar{m}_A$ is a monomorphism. Thus $\bar{A}$, as a subgroup of the $\mathbb{Q}(P)$-module CA is P-torsion-free. To show that $\bar{A} \xrightarrow{\bar{c}} A$ is a weak $\mathtt{Torf}(P)$-coreflection, let $B \xrightarrow{f} A$ be a homomorphism with P-torsion-free domain. Since $B \overset{m_B}{\hookrightarrow} DB$ is an embedding and DA is injective there exists a homomorphism $DB \xrightarrow{g} DA$ with $g \circ m_B = m_A \circ f$. Since, by Lemma 4.5, DB is P–torsion–free there exists a homomorphism
$DB \xrightarrow{h} CA$ with $g = c_A \circ h$. Thus, by the pullback property, there exists a homomorphism $B \xrightarrow{k} \bar{A}$ such that the following diagram commutes:

$$\begin{tikzcd}
B \arrow[rrr, hook, "m_B"] \arrow[dr, "k"] \arrow[ddr, "f"'] & & & DB \arrow[dl, "h"'] \arrow[ddl, "g"] \\
& \bar{A} \arrow[r, hook, "\bar{m}_A"] \arrow[d, "\bar{c}_A"] & CA \arrow[d, "c_A"'] & \\
& A \arrow[r, hook, "m_A"'] & DA &
\end{tikzcd}$$

Consequently $\bar{A} \xrightarrow{\bar{c}_A} A$ is a weak $\mathtt{Torf}(P)$-coreflection. Since $\mathtt{Torf}(P)$ is closed under the formation of retracts, it is almost coreflective in $\mathtt{Ab}$.

That $\mathtt{Torf}(P)$, for $P \neq \emptyset$, is not coreflective in $\mathtt{Ab}$ follows from the fact that $\mathtt{Torf}(P)$ is not closed under the formation of coequalizers in $\mathtt{Ab}$. Consider, e.g., for some $p \in P$ the zero map $f \colon \mathbb{Z} \longrightarrow \mathbb{Z}$ and the map $g \colon \mathbb{Z} \longrightarrow \mathbb{Z}$, defined by $g(n) = pn$. Then the natural quotient map $\mathbb{Z} \longrightarrow \mathbb{Z}_p$ is a coequalizer of f and g, but whereas $\mathbb{Z}$ belongs to $\mathtt{Torf}(P)$, the group $\mathbb{Z}_p$ fails to do so. ■

Remark 4.8 The above mentioned duality carries even further in case P is the set of all primes (see Section 5). However, in all other cases, it is more limited; e.g., $\mathtt{Torf}(P)$ fails to be the epireflective hull of $\mathbb{Q}(P)$, since $\mathbb{Q}$ is P-torsion-free but the only homomorphism from $\mathbb{Q}$ to $\mathbb{Q}(P)$ is the zero-morphism.

4.2 A Curious Special Case

If we specialize the concepts from Section 3 to the case where P consists of all prime numbers, a curious phenomenon occurs. In this case P will be omitted in our notation from Definition 4.1, so that $\mathcal{Q}$ is the subcategory of $\mathtt{Ab}$, consisting of all torsion-free, divisible groups, i.e., (up to isomorphism) the category of $\mathbb{Q}$-modules, $\mathtt{Torf}$ is the subcategory of $\mathtt{Ab}$, consisting of all torsion-free groups, and $\mathtt{Div}$ is the subcategory of $\mathtt{Ab}$, consisting of all divisible groups.

Proposition 4.8 $\mathtt{Torf}$ *is the epireflective hull of* $\mathbb{Q}$ *in* $\mathtt{Ab}$.

Proof. The epireflective hull $\mathcal{H}$ of $\mathbb{Q}$ in $\mathtt{Ab}$ consists precisely of those groups that are isomorphic to a subgroup of some power $\mathbb{Q}^I$ of $\mathbb{Q}$. Obviously $\mathcal{H} \subset \mathtt{Torf}$.

Let A be a torsion-free group and let a be a non-zero element of A. The unique homomorphism $f \colon \mathbb{Z} \longrightarrow A$ with $f(1) = a$ is an embedding. Since $\mathbb{Q}$ is injective in $\mathtt{Ab}$ there exists a homomorphism $g \colon A \longrightarrow \mathbb{Q}$ such that the diagram

$$\begin{tikzcd} \mathbb{Z} \arrow[rr, hook, "f"] \arrow[dr, "j"'] & & A \\ & \mathbb{Q} \arrow[ur, dashed, "g"'] & \end{tikzcd}$$

commutes, where j denotes the inclusion map. Then $g(a) = 1$. Consequently

$$\bigcap\{\ker g \mid g \in \hom(A, \mathbb{Q})\} = \{0\}.$$

Thus the kernel of the canonical map

$$A \longrightarrow \mathbb{Q}^{\hom(A,\mathbb{Q})}$$

is zero. Thus A belongs to $\mathcal{H}$. The proof is complete. ■

Proposition 4.9 `Div` *is the monocoreflective hull of* $\mathbb{Q}$ *in* `Ab`.

Proof. Immediate from Corollary 4.6. We add a more direct proof: The monocoreflective hull $\mathcal{H}$ of $\mathbb{Q}$ in `Ab` consists precisely of the homomorphic image of copowers ${}^{I}\mathbb{Q}$ ($=$ sums $\mathbb{Q}^{(I)}$) of $\mathbb{Q}$. Obviously $\mathcal{H} \subset$ `Div`.

Let A be a divisible group and let a be an element of A. Let $f\colon \mathbb{Z} \longrightarrow A$ be the unique homomorphism with $f(1) = a$, and let $j\colon \mathbb{Z} \hookrightarrow \mathbb{Q}$ be the inclusion map. As a divisible group, A is injective. Thus there exists a homomorphism $g\colon \mathbb{Q} \longrightarrow A$ such that the diagram

$$\begin{tikzcd} \mathbb{Z} \arrow[rr, hook, "j"] \arrow[dr, "f"'] & & \mathbb{Q} \arrow[dl, dashed, "g"] \\ & A & \end{tikzcd}$$

commutes. Then $g(1) = a$. Consequently the canonicali map ${}^{\hom(\mathbb{Q},A)}\mathbb{Q} \longrightarrow A$ is surjective. Thus A belongs to $\mathcal{H}$. The proof is complete. ■

Theorem 4.10 *(1)* $\mathbb{Q}$ *is the reflective hull of* $\mathbb{Q}$ *in* `Ab`*,*

(2) $\mathbb{Q}$ *is the coreflective hull of* $\mathbb{Q}$ *in* `Ab`*,*

(3) $\mathbb{Q}$ *is the intersection of the epireflective hull of* $\mathbb{Q}$ *in* `Ab` *and the monocoreflective hull of* $\mathbb{Q}$ *in* `Ab`*,*

Proof. The result follows immediately from Theorem 2.1, Propositions 4.8 and 4.9, and the following simple and known facts:

The category $\mathcal{Q}$ is closed under the formation of products, coproducts and equalizers in `Ab`, and each of its objects is isomorphic to a copower pf $\mathbb{Q}$. In $\mathbb{Q}$ –**Vec** every object is an equalizer of a suitable pair of morphisms $\mathbb{Q}^I \overset{r}{\underset{s}{\rightrightarrows}} \mathbb{Q}^K$ between powers of $\mathbb{Q}$. ■

References

[1] J. Adámek, H. Herrlich, and G. E. Strecker. Abstract and Concrete Categories, Wiley, 1990.

[2] L. Fuchs. Infinite Abelian Groups I, Acad. Press, 1970.

[3] E. Giuli. Two families of cartesian closed topological categories, *Quaestiones Math.*, **6** (1983), 353–362.

[4] H. Herrlich. Universal topology, in *Categorical Topology*, H.L. Bentley, H. Herrlich M. Rajagopalan and H. Wolff, eds., Heldermann Verlag, 1984, 223–281.

[5] H. Herrlich and R. Lowen On simultaneously reflective and coreflective subconstructs, *Proc. Symposium on Categorical Topology*, Rondebosch, 1994, 121–129, Univ. Cape Town, Rondebosch, 1999.

[6] M. Hušek. Reflective and coreflective subcategories of UNIF and TOP, in *Seminar Uniform Spaces*, Z. Frolík, ed., Prague 1973, 113–126.

[7] M. Hušek. Lattices of reflections and coreflections in continuous structures, *Springer Lecture Notes Math.*, **540** (1976), 404–424.

[8] V. Kannan. Reflexive cum coreflexive subcategories in topology, *Math. Annalen*, **195** (1972), 168–174.

[9] S. Mac Lane. *Categories for the Working Mathematician*, Springer–Verlag, 1971.

Robert El Bashir and Miroslav Hušek
Department of Mathematics
Charles University
Sokolovská 83, Prague 18675, Czech Republic
bashir@karlin.mff.cuni.cz
mhusek@karlin.mff.cuni.cz

Horst Herrlich
Department of Mathematics
University of Bremen
Bremen, Germany
herrlich@math.uni-bremen.de